Challenges in Dynamic Macroeconomics

Dissertation

for the

Faculty of Economics, Business Administration and Information Technology

of the

University of Zurich

to achieve the title of

Doctor of Philosophy

in Economics

presented by

Timo Boppart

from Erlenbach ZH

approved in October 2012 at the request of

Prof. Dr. Dr. Josef Falkinger

Prof. Dr. Fabrizio Zilibotti

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorises the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, on the 24th of October, 2012

Chairman of the Doctoral Committee: Prof. Dr. Dieter Pfaff

Acknowledgments

While writing this dissertation I worked as a research assistant for Josef Falkinger at the Chair of Public Finance and Macroeconomics at the University of Zurich. This is the place where I have learned to do rigorous economic research and I am very grateful to my supervisor Josef Falkinger and the current and former members of his group for this excellent time. I am deeply indebted to Josef Falkinger for his support, guidance, encouragement and trust.

I am very grateful to Fabrizio Zilibotti for his innumerable positive spillovers on the entire Department of Economics and specifically on my work and development. I am very happy that Fabrizio Zilibotti agreed to co-advice this dissertation. Moreover, I would also like to thank Josef Zweimüller – who gave me my first lesson of Macroeconomics back in 2002 – for a lot of illuminating discussions over the years.

The Department of Economics of the University of Zurich is a very inspiring and constructive research environment and I am very thankful to many fellow PhD students, post docs, co-authors, members of the faculty as well as guests who contributed to this great atmosphere and helped to improve my work during seminars and bilateral discussions.

Finally, I wish to thank my family and friends who have believed in me and gave me their support during these years.

Zurich, July 2012 Timo Boppart

Chapter 1

Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences

Chapter Summary

Growth is associated with (i) shifts in the sectoral structure of the economy, (ii) changes in relative prices and (iii) the Kaldor facts. Moreover, (iv) cross-sectional data shows systematic differences in the expenditure structure. This paper presents a growth model which is consistent with (i)-(iv) at the same time, a result the existing literature has not been able to generate. The theory is simple and parsimonious and contains an analytical solution. The model's functional form and the cross-sectional data are exploited to estimate the relative importance of price and income effects as determinants of the structural change.

1.1 Introduction

It is a well documented empirical fact that economic growth is associated with significant shifts in the sectoral output, employment and consumption structure (see e.g. Kuznets, 1957 and Kongsamut, Rebelo and Xie, 2001). This phenomenon is summarized under the term "structural change". As an example, Figure 1.1 shows the relative decline of the goods sector (or the rise of the service sector) in the U.S. after World War II. On a logarithmic scale the evolution of the expenditure share devoted to goods is well approximated by a linear downward sloping trend (see dashed line). The slope of this linear fit suggests that the expenditure share devoted to goods decreases (on average) at a constant annualized rate of one percent.

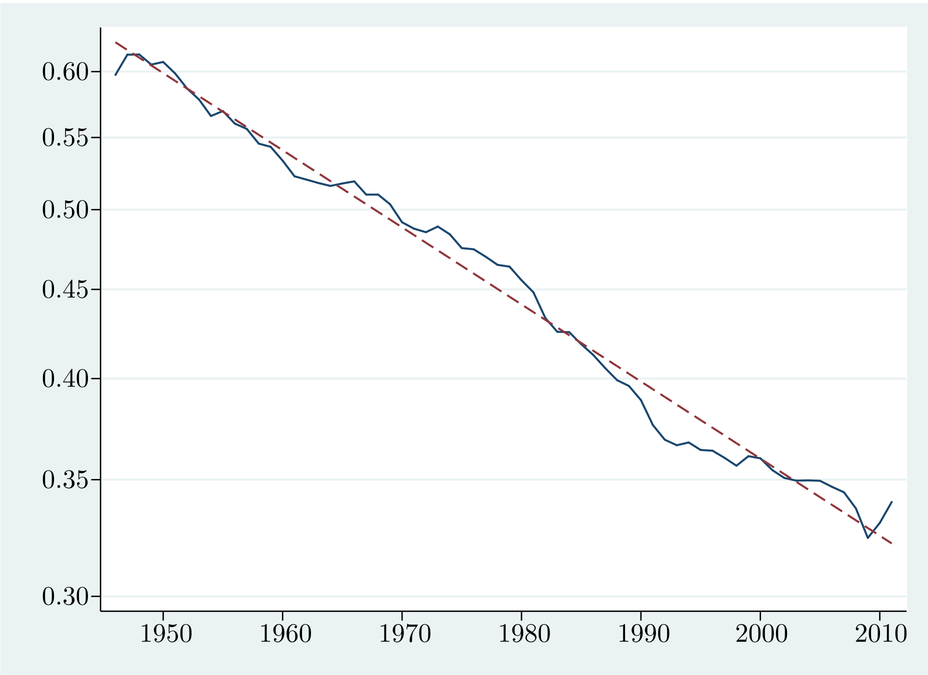


Figure 1.1: Expenditure share of goods

Notes: The figure plots the share of personal consumption expenditures devoted to goods in the U.S. on a logarithmic scale. The dashed line represents the predicted values obtained by regressing the logarithmized expenditure share on time and a constant. The estimated slope coefficient and its standard error is −0.0102 and 0.00015, respectively. The regression attains an R^2 of 0.986. Source: BEA, NIPA table 1.1.5.

The nonbalanced nature of growth is displayed in prices too. Figure 1.2 plots the evolution of the relative consumer price between goods and services on a logarithmic scale. Apart from the two oil crises in 1973 and 1979, the series is fairly good approximated by a constant annualized growth rate of -1.6 percent (see dashed line).

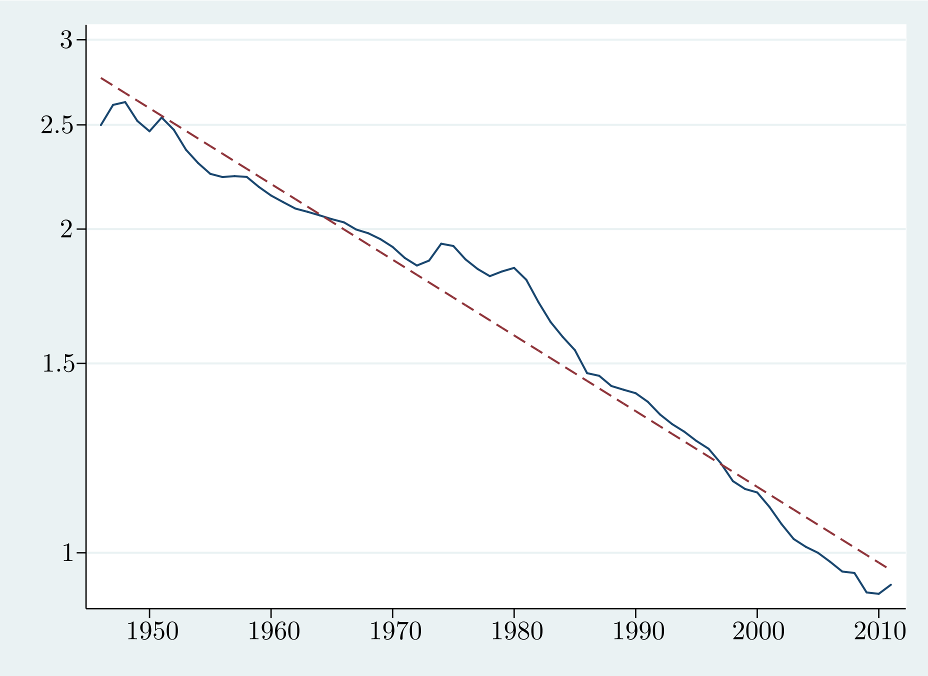


Figure 1.2: Relative price between goods and services

Notes: The figure plots the relative consumer price between goods and services on a logarithmic scale. The dashed line represents the predicted values obtained by regressing the logarithmized relative price on a constant and time. The estimated slope coefficient and its standard error is -0.0162 and 0.00037, respectively. The regression attains an R^2 of 0.968. Source: BEA, NIPA table 1.1.4.

Beyond the nonbalanced characteristics at the sectoral level, aggregate variables present a balanced picture of growth. Actually, the post-war U.S. often serves as a prime example of balanced growth on the aggregate. Balanced growth is best summarized by the Kaldor facts. These stylized facts state that the growth rate of real per-capita output, the real interest rate, the capital-output ratio and the labor income share are constant over time (see Kaldor, 1961). As a consequence, comprehensive models of structural change should also replicate the Kaldor facts.

The paper by Ngai and Pissarides (2007) reconciles structural change, relative price dynamics and the Kaldor facts in a growth model with endogenous savings. Another paper that emphasizes relative price dynamics as a driver of structural change is the one by Acemoglu and Guerrieri (2008).[1] Both theoretical models – Ngai and Pissarides (2007) and Acemoglu and Guerrieri (2008) – feature a constant elasticity of substitution across sectors. However, the relative nominal expenditures of goods declined in the U.S. at a faster rate than the relative price of goods. Hence, with relative price effects alone, theories with a constant elasticity of substitution cannot replicate the observed structural change quantitatively.[2]

Acemoglu and Guerrieri (2008) emphasize that income effects are an "undoubtly important" determinant of structural change. Nevertheless, both Ngai and Pissarides (2007) and Acemoglu and Guerrieri (2008) abstract from non-homotheticity of preferences.[3] Empirically, there is clear evidence for an income effect. Figure 1.3 plots the expenditure shares devoted to goods for the different income quintiles. Rich households exhibit a significantly lower expenditure share of goods then poor households. Moreover, on a logarithmic scale, the expenditure shares in Figure 1.3 are parallel and decline linearly. This suggests that expenditure shares devoted to goods of rich and poor households decline at the same (constant) growth rate as the aggregate series. With non-unitary expenditure elasticities of demand, increases in real per-capita expenditure levels (due to growth) affect the sectoral expenditure shares.[4] Kongsamut, Rebelo and Xie (2001) and Foellmi and Zweimueller (2008) reconcile non-homothetic preferences and the Kaldor facts in an otherwise standard growth models with intertemporal optimization. However, in order to obtain balanced aggregate growth, both theories have to exclude relative price effects.[5] Hence, as pointed out by Buera and Kaboski (2009a), none of the existing models with endogenous savings and balanced aggregate growth, allows us to discuss both forces of structural change - relative price and income effects.

The contributions of this paper are as follows: First, it presents a neoclassical growth theory with intertemporal optimization, which reconciles the Kaldor facts with structural change simultaneously determined by relative price and income effects. By postulating non-Gorman preferences the paper

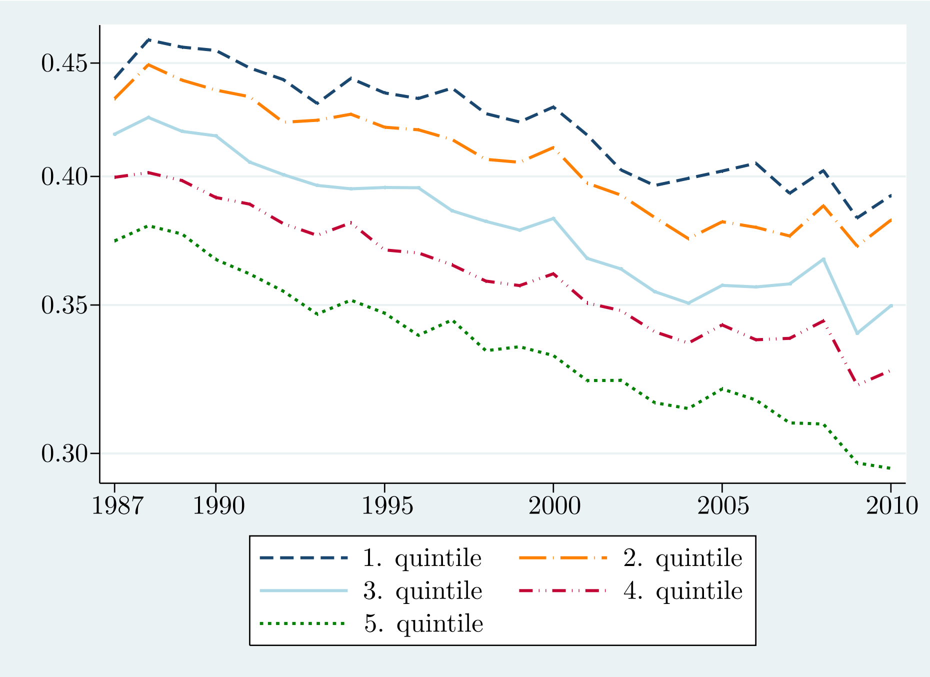


Figure 1.3: Cross-sectional variation in expenditure structure

Notes: The figure plots the expenditure share devoted to goods for each income quintile of the U.S. on a logarithmic scale. The following expenditure categories are considered as services: food away from home; shelter; utilities, fuels, and public services; other vehicle expenses; public transportation; health care; personal care; education; cash contributions; personal insurance and pensions. The remaining categories are considered as goods. The sample consists of expenditure data of 450,602 quarters (and 165,887 households). Observations with missing income reports, with non-positive food expenditures or with an expenditure share of goods outside $[0,1]$ have been excluded. The quintiles refer to total household after tax labor earnings plus transfers per OECD-modified equivalence scale. If we observe for a household more than one income report, the income data of the year in which the expenditure quarter lies is taken. Source: Consumer Expenditure Survey.

also illustrates a tractable (dynamic) framework which allows for effects of inequality on the aggregate demand structure. Second, the paper illustrates that the theory can replicate the shape and magnitude of structural change and relative price dynamic identified in Figure 1.1 and 1.2. Moreover, the model is consistent with cross-sectional expenditure structure differences and the parallel evolution of logarithmized expenditure shares of different income groups, depicted in Figure 1.3. Finally, a structural estimation allows us to decompose the structural change into an income and substitution effect.[6]

The paper consists of four sections: Section 1.2 presents the theoretical growth model. In section 1.3 an estimation of the relative importance of income and substitution effects as determinants of structural change is carried out. Finally, section 1.4 concludes.

1.2 Theoretical model

There is a unit interval of (heterogeneous) households indexed by $i \in [0, 1]$. Each household consists of $N(t)$ identical members, where $N(t)$ grows at an exogenous rate $n \geq 0$. $N(0)$ is normalized to one, so we have $N(t) = \exp[nt]$. Each member of household i is endowed with $l_i \in (\bar{l}, \infty)$, $\bar{l} > 0$, units of labor and $a_i(0) \in [0, \infty)$ units of initial wealth. These per-capita factor endowments can differ across households. Labor is supplied inelastically at every instant of time. Consequently, the aggregate labor supply $L(t) \equiv N(t) \int_0^1 l_i \, di$, grows at constant rate n.

1.2.1 Preferences

All households have the following additively separable representation of intertemporal preferences

$$\mathcal{U}_i(0) = \int_0^\infty \exp\left[-(\rho - n)t\right] V\left(P_1(t), P_2(t), e_i(t)\right) dt, \tag{1.1}$$

where $\rho \in (n, \infty)$ is the rate of time preference and $V(P_1(t), P_2(t), e_i(t))$ is an indirect instantaneous utility function of each household member. This instantaneous utility function is specified over the prices of the two consumption goods, $P_1(t)$ and $P_2(t)$, and the nominal per-capita expenditure level of household i, $e_i(t)$. Henceforth, the first consumption good is called "good", whereas the second consumption good is "service". The indirect instantaneous utility function takes the following form

$$V(P_1(t), P_2(t), e_i(t)) = \frac{1}{\epsilon}\left[\frac{e_i(t)}{P_2(t)}\right]^{\epsilon} - \frac{\beta}{\gamma}\left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma} - \frac{1}{\epsilon} + \frac{\beta}{\gamma}, \tag{1.2}$$

where $0 \leq \epsilon \leq \gamma < 1$ and $\beta, \gamma > 0$.[7] It will be shown below that these preferences imply a household behavior which is consistent with the facts emphasized in the introduction.[8] The specified intratemporal utility function represents a subclass of "price independent generalized linearity" (PIGL) preferences defined by Muellbauer (1975) and Muellbauer (1976). The PIGL class of preferences is more general than the Gorman class. Nevertheless, PIGL preferences avoid an aggregation problem. Expenditure shares of the aggregate economy coincide with those of a household with a "representative" expenditure level (the representative household in Muellbauer's sense). Moreover, PIGL preferences ensure that this representative expenditure level is independent of prices. Because Engel curves are patently non-linear, PIGL preferences have explicitly an empirical justification and are widely used in expenditure system estimations (see e.g. the "Quadratic Expenditure System" (QES) by Howe, Pollak and Wales, 1979 or the "Almost Ideal Demand System" (AIDS) by Deaton and Muellbauer, 1980).
Lemma 1.1 shows that function (1.2) satisfies the standard properties of a utility function.

LEMMA 1.1. *Function* (1.2),

(i) is a valid indirect utility specification that represents a preference rela-

tion defined over goods and services if and only if

$$e_i(t)^\epsilon \geq \left[\frac{1-\epsilon}{1-\gamma}\right] \beta P_1(t)^\gamma P_2(t)^{\epsilon-\gamma}, \tag{1.3}$$

(ii) is increasing and strictly concave in $e_i(t)$.

Proof. See Appendix A.1.1. Q.E.D.

Henceforth, I assume that condition (1.3) is fulfilled. Later, two conditions in terms of exogenous parameters are stated, which jointly ensure condition (1.3) for all individuals, at each date. Strict concavity of the intratemporal utility function is a necessary condition for intertemporal optimization, which will be addressed below.

The characteristics of the intratemporal preferences are best discussed in terms of the associated expenditure system. Applying Roy's identity, we get the following lemma.

LEMMA 1.2. *At each point in time, intratemporal preferences imply the following expenditure system*

$$x_1^i(t) = \beta \frac{e_i(t)}{P_1(t)} \left[\frac{P_2(t)}{e_i(t)}\right]^\epsilon \left[\frac{P_1(t)}{P_2(t)}\right]^\gamma, \tag{1.4}$$

and

$$x_2^i(t) = \frac{e_i(t)}{P_2(t)} \left[1 - \beta \left[\frac{P_2(t)}{e_i(t)}\right]^\epsilon \left[\frac{P_1(t)}{P_2(t)}\right]^\gamma\right], \tag{1.5}$$

where $x_j^i(t)$, $j = 1, 2$, is household i's per-capita consumption of goods/services at date t.

The expenditure system reveals, that the demand for goods, $x_1^i(t)$, is an exponential function of order $1-\epsilon$ of the per-capita expenditure level. The expenditure shares devoted to the two consumption sectors, $s_j^i(t)$; $j = 1, 2$, can be expressed as

$$s_1^i(t) = \beta \left[\frac{P_2(t)}{e_i(t)}\right]^\epsilon \left[\frac{P_1(t)}{P_2(t)}\right]^\gamma \text{ and } s_2^i(t) = 1 - \beta \left[\frac{P_2(t)}{e_i(t)}\right]^\epsilon \left[\frac{P_1(t)}{P_2(t)}\right]^\gamma. \tag{1.6}$$

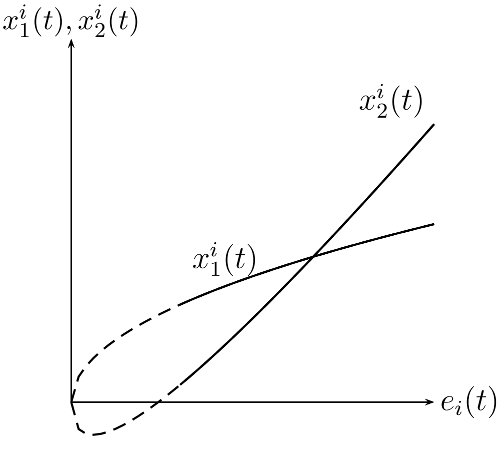


Figure 1.4: Engel curves

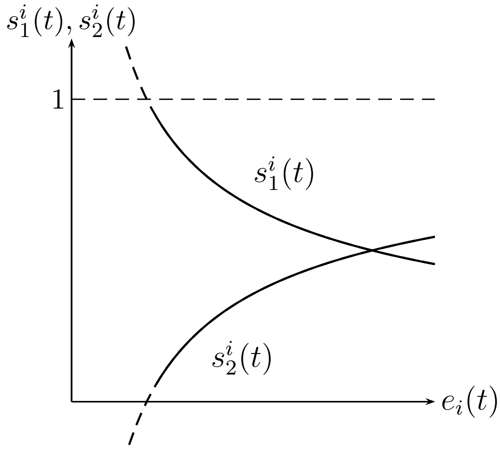


Figure 1.5: Expenditure shares

Notes: As indicated by the dashed sections, preferences are only well defined, if condition (1.3) holds (i.e. $e_i(t)$ exceeds a certain threshold).

For $\epsilon > 0$, Figure 1.4 and 1.5 plot the Engel curves and the sectoral expenditure shares as a function of the per-capita expenditure level. In general, as the non-linear Engel curves reveal, preferences are non-homothetic and even do not fall into the Gorman class.

The elasticity of substitution across sectors and the expenditure elasticities of demand control the magnitude and direction of the income and substitution effects on expenditure shares. Growing real per-capita expenditure levels generate – according to the income effect – an increasing expenditure share of the sector, whose expenditure elasticity of demand exceeds unity. Besides, the substitution effect implies that if the elasticity of substitution is strictly less than unity the sector which experiences a relative price increase, gains in terms of expenditure shares. If the elasticity of substitution is larger then one, the structural change would run in the opposite direction. The next lemma characterizes the two important elasticities.

LEMMA 1.3. *The intratemporal preferences,* (1.2)*, imply that* *(i) the elasticity of substitution between goods and services,*

$$\sigma_i(t) = 1 - \gamma - \frac{\beta \left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma}}{\left[\frac{e_i(\cdot)}{P_2(t)}\right]^{\epsilon} - \beta \left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma}} \left[\gamma - \epsilon\right], \tag{1.7}$$

is strictly less than unity (for all households at each date).

(ii) with $\epsilon > 0$, the expenditure elasticity of demand is positive, but strictly smaller than one for goods and larger than one for services.

(iii) with $\epsilon = 0$ we have homothetic preferences (expenditure elasticities of both sectors are equal to unity).

Proof. The Allen-Uzawa formula for the elasticity of substitution reads $\sigma_i(t) = \frac{\partial x_1^{i,H}(t)}{\partial P_2(t)} \frac{e_i(t)}{x_1^{i,H}(t) x_2^{i,H}(t)}$, where $x_j^{i,H}(t)$ is the Hicksian per-capita demand of household i for sector $j = 1, 2$. Plugging in the expressions for the Hicksian demand, simplifying and substituting $V_i(t)$ by (1.2), we obtain (1.7). With $\gamma > 0$ and (1.3), $\sigma_i(t)$ is strictly smaller than one since $\gamma \geq \epsilon$. This completes part (i). Part (ii) and (iii) follow immediately from (1.4) and (1.5). Q.E.D.

Several things are worth noting: First, the restrictions on the preference parameters ϵ and γ are such that the elasticity of substitution is strictly less than unity. In the literature there seems to be a consensus that this is the empirically relevant case.[9] This notion is also confirmed in section 1.3.

Second, in general, the elasticity of substitution varies over time and across households. Nevertheless, there is a special case with $\gamma = \epsilon$, in which the elasticity of substitution is constant for all households at each date.

Third, with $\epsilon = 0$, we have homothetic preferences and consequently no income effect on expenditure shares. In contrast, as long as $\epsilon > 0$, goods are necessities with an expenditure elasticity of demand strictly smaller than one.[10]

Next, we turn to the household's intertemporal optimization problem. Households maximize (1.1) with respect to $\{e_i(t), a_i(t)\}_{t=0}^{\infty}$, subject to the budget constraint

$$\dot{a}_i(t) = [r(t) - n]\, a_i(t) + w(t)l_i - e_i(t), \tag{1.8}$$

and a standard transversality condition, which can be expressed as

$$\lim_{t\to\infty} e_i(t)^{\epsilon-1} P_2(t)^{-\epsilon} a_i(t) \exp\left[-(\rho - n)t\right] = 0. \tag{1.9}$$

$r(t)$ and $w(t)$ is the (nominal) interest and wage rate, respectively, and $a_i(t)$ denotes the per-capita wealth of household i at date t. $a_i(0)$ is exogenously given. The result of intertemporal household optimization is summarized in the next lemma.

LEMMA 1.4. *Intertemporal optimization yields the Euler equation*

$$(1 - \epsilon) g_{e_i}(t) + \epsilon g_{P_2}(t) = r(t) - \rho, \tag{1.10}$$

where $g_{e_i}(t)$ is the growth rate of per-capita consumption expenditures of household i and $g_{P_2}(t)$ is the growth rate of the price of services at date t.

Proof. The current value Hamiltonian of the household's intertemporal optimization is given by $\mathcal{H} = V(\cdot) + \lambda_i(t)\left[a_i(t)\left[r(t) - n\right] + w(t)l_i - e_i(t)\right]$. We can then derive the first-order conditions $\dot{\lambda}_i(t) = \lambda_i(t)\left[\rho - r(t)\right]$ and $e_i(t)^{\epsilon-1} P_2(t)^{-\epsilon} = \lambda_i(t)$, which can be rewritten as (1.10). Q.E.D.

The Euler equation takes the same functional form as in the standard neoclassical growth model with CRRA preferences. Additionally, since $g_{e_i}(t)$ is the only term that involves a household index i, the Euler equation implies that the growth rate of the per-capita expenditure levels is the same for all households at a given point in time, or formally,

$$g_{e_i}(t) = g_e(t),\ \forall i. \tag{1.11}$$

Together with the desirable aggregation properties specific to all PIGL preferences, the feature that all expenditure levels grow pari passu, simplifies the equilibrium analysis dramatically. Let us define $E(t)$ as the aggregate consumption expenditures and $X_j(t)$ as the aggregate demand for consumption $j = 1, 2$ at date t (i.e. $E(t) \equiv N(t) \int_0^1 e_i(t)di$ and $X_j(t) \equiv N(t) \int_0^1 x_j^i(t)di$, $j = 1, 2$). Then, household behavior is summarized by the following proposition.

PROPOSITION 1.1. *Under consumer optimization,*

(i) the intertemporal behavior of the demand side is fully characterized by the following Euler equation, budget constraints and transversality conditions:

$$(1-\epsilon)\left[g_E(t) - n\right] + \epsilon g_{P_2}(t) = r(t) - \rho, \ \forall t, \tag{1.12}$$

where $g_E(t)$ is the growth rate of $E(t)$,

$$\dot{a}_i(t) = \left[r(t) - n\right] a_i(t) + w(t)l_i - e_i(0) \exp\left[\int_0^t g_E(\varsigma) - n \, d\varsigma\right], \ \forall i, t, \tag{1.13}$$

and

$$\lim_{t\to\infty} a_i(t) \exp\left[-\int_0^t r(\varsigma) - n \, d\varsigma\right] = 0, \ \forall i, \tag{1.14}$$

where $a_i(0)$, $\forall i$, is exogenously given.

(ii) the aggregate expenditure share devoted to goods, $S_1(t) \equiv \frac{P_1(t)X_1(t)}{E(t)}$, is given by

$$S_1(t) = \beta \left[\frac{P_2(t)}{\frac{E(t)}{N(t)}}\right]^{\epsilon} \left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma} \phi, \tag{1.15}$$

where $\phi \equiv \int_0^1 \left[\frac{e_i(0)N(0)}{E(0)}\right]^{1-\epsilon} di$ is a scale invariant (inverse) measurement of inequality of per-capita consumption expenditures across households. Furthermore, we have

$$E(t) = P_1(t)X_1(t) + P_2(t)X_2(t). \tag{1.16}$$

(iii) a household with $e_i(t) = \frac{E(t)}{N(t)}\phi^{-\frac{1}{\epsilon}} \equiv e^{RA}(t)$ *is the representative agent in Muellbauer's sense.*[11]

Proof. (1.11) implies $g_{e_i}(t) = g_E(t) - n$, $\forall i$, allowing us to rewrite (1.10) as (1.12). Substituting $e_i(t)$ in (1.8) by $e_i(0)\exp\left[\int_0^t g_E(\varsigma) - n\, d\varsigma\right]$ yields (1.13). Using (1.10) in (1.9) and ignoring the positive constant $e_i(0)$ gives (1.14). Aggregation of individual demands gives

$$X_1(t) = \beta P_1(t)^{-1} P_2(t)^{\epsilon}\left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma}\left[\frac{E(t)}{N(t)}\right]^{-\epsilon} E(t)\phi(t),$$

$$X_2(t) = \frac{E(t)}{P_2(t)} - \beta P_2(t)^{\epsilon-1}\left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma}\left[\frac{E(t)}{N(t)}\right]^{-\epsilon} E(t)\phi(t),$$

where $\phi(t) = \int_0^1 \left[\frac{e_i(t)N(t)}{E(t)}\right]^{1-\epsilon} di$. These two equations imply (1.15) and (1.16), where $\phi(t)$ is constant over time because of (1.11) and because it is scale invariant in all $e_i(t)$. For part (iii): (1.6) and (1.15) show that a household exhibits the same expenditure shares as the aggregate economy if $e_i(t) = \frac{E(t)}{N(t)}\phi^{-\frac{1}{\epsilon}}$. Q.E.D.

This proposition fully characterizes the demand side of this economy. Given a path of production factor, good and service prices, $\{r(t), w(t), P_1(t), P_2(t)\}_{t=0}^{\infty}$, equations (1.12)-(1.16) define the equilibrium evolution of the level and structure of aggregate consumption expenditures. Since in general, the intratemporal preferences do not fall into the Gorman class, a representative agent in the narrower sense does not apply and the distribution of per-capita expenditure levels matters. Nevertheless, the tractability of the specified preferences allows us to write the aggregate demand of goods and services as a function of just two terms: the aggregate expenditure level, $E(t)$, and a summary statistic of the distribution of per-capita expenditure levels at date $t = 0$, denoted by ϕ. This is the outcome of two special properties:
First, the fact that preferences are part of the "generalized linearity" class, allows for a representative agent in Muellbauer's sense (see Muellbauer, 1975 and Muellbauer, 1976). A household with the representative expenditure

level, $e^{RA}(t)$, exhibits the same expenditure shares as the aggregate economy. Moreover, since preferences are even part of the PIGL class, the representative expenditure level is independent of prices. Consequently, aggregate demand can be expressed as a function of $E(t)$ and the scale invariant inequality measure of per-capita expenditure levels at date t, $\phi(t) = \int_0^1 \left[\frac{e_i(t)N(t)}{E(t)}\right]^{1-\epsilon} di$.

The second property is that intertemporal optimization implies for all households the same per-capita expenditure growth rate at any given point in time (see (1.11)). Then, $\phi(t)$ is constant over time and can therefore be expressed as a function of the $e_i(0)$ distribution.[12] This tractability allows me to solve the model analytically, despite household heterogeneity, non-Gorman intratemporal preferences and intertemporal optimization.[13]

ϕ can be related to an Atkinson index of expenditure inequality. To see this, note that the Atkinson index (Atkinson, 1970) is defined as

$$I_A\left(\zeta, \{e_i(t)\}_{i=0}^1\right) = 1 - \frac{N(t)}{E(t)}\left[\int_0^1 e_i(t)^{1-\zeta} di\right]^{\frac{1}{1-\zeta}},$$

with the parameter $\zeta \geq 0$ being the relative inequality aversion. Then, we can write

$$\phi(t) = \left[1 - I_A\left(\epsilon, \{e_i(t)\}_{i=0}^1\right)\right]^{1-\epsilon},$$

i.e. ϕ is a negative, monotonic transformation of the Atkinson inequality index with $\zeta = \epsilon$. Hence, ϕ is an ordinally equivalent of the inverse of an Atkinson index. This justifies our interpretation of ϕ as an inverse measurement of expenditure inequality fulfilling the principle of transfers, scale invariance and decomposability (see Cowell, 2000).

To close the model, i.e. in order to determine the equilibrium path of production factor, good and service prices, the production side of the economy remains to be specified.

1.2.2 Production

There are three output goods: the output of the two consumption sectors $Y_1(t)$ and $Y_2(t)$ and an "investment good", $Y_3(t)$, which can be transformed one-to-one into capital, $K(t)$. Capital depreciates at constant rate $\delta \geq 0$. This implies for the law of motion of capital

$$\dot{K}(t) = X_3(t) - \delta K(t), \tag{1.17}$$

where $X_3(t)$ is aggregate gross investment (in terms of investment goods) at date t. The consumption sectors produce under perfect competition according to the following technologies

$$Y_j(t) = \exp\left[g_j t\right] L_j(t)^{\alpha} K_j(t)^{1-\alpha},\ j = 1, 2, \tag{1.18}$$

where $L_j(t)$ and $K_j(t)$ denotes labor and capital, respectively, allocated to sector j at date t. Both production factors are fully mobile across sectors. $\alpha \in (0, 1)$ is the output elasticity of labor, which is identical across sectors. Total factor productivity (TFP) expands at a constant, exogenous, sector-specific rate $g_j \geq 0$.[14] The investment good is produced by a linear technology

$$Y_3(t) = AK_3(t), \tag{1.19}$$

with $A > \delta$. The market of investment goods is competitive, too. Henceforth, I normalize the price of the investment good at each date to one, i.e. $P_3(t) = 1$, $\forall t$. The production side of this economy is similar to the one in Rebelo (1991).[15] $K(t)$ is a "core" capital good, whose production does not involve nonreproducible factors. This makes endogenous growth feasible. But as long as $g_j \neq 0$, for some $j = 1, 2$, the economy also consists of an exogenous driver of growth.

It is worthwhile to discuss shortly in which respects the functional forms of the production functions can be generalized. First, the AK structure of the investment good sector is not essential. It can be relaxed to any neoclassical production function with constant Harrod-neutral productivity growth, i.e. $Y_3 = F(K_3(t), \exp[g_3 t] L_3(t))$. With this more general specification transitional dynamics arise along which capital per effective labor, $\frac{K(t)}{\exp[g_3 t] L(t)}$, adjusts. On the aggregate this transition is identical to the one in a standard one-sector neoclassical growth model. And in the steady state the equilibrium looks as the one with the AK technology and the Kaldor facts hold. So the AK technology allows us to focus more directly on the main dynamics: the coexistence of structural change and balanced growth on the aggregate.

The production functions of the consumption sectors must ensure along the equilibrium path the following two properties: (i) For the consumption sectors, the overall labor income share must be constant and (ii) the relative price between services and the investment good, $\frac{P_2(t)}{P_3(t)}$, must change at a constant rate. Requirement (i) is common to all structural change models aiming to be consistent with the Kaldor facts. It is typically accommodated by a *constant* and *identical* steady state labor income share in both sectors. This can either be achieved by assuming that the production functions of sector 1 and 2 are – up to a time varying Hicks-neutral productivity term – identical to the one of the investment good, i.e. $Y_j(t) = A_j(t) F(K_j(t), \exp[g_3 t] L_j(t))$, $j = 1, 2$.[16] Alternatively, the production technologies may differ from the one of the investment good. But then we need, up to a time varying productivity term, $A_j(t)$, *identical* Cobb-Douglas technologies in both consumption sectors $j = 1, 2$. This is the specification chosen above (and also in Ngai and Pissarides, 2007). Requirement (ii) is specific to this model and implies that the time varying productivity term of the service sector must grow at a constant rate, i.e. $A_2(t) = A_2(0) \exp[g_2 t]$.[17]

Finally, it is worth noting that the entire model is specified in terms of final output as opposed to value added. This means that in order to derive theoretical implications for sectoral value added shares the exact production processes with intermediate inputs have to be specified (see Herrendorf, Rogerson and Valentinyi, 2009 for the empirical differences of these two perspectives). In this light the assumption of identical capital intensity of the good and service sector seems not unrealistic. Valentinyi and Herrendorf (2008) estimate labor income shares for gross manufacturing output, gross service output, overall consumption and total gross output that are all between 0.65 and 0.67. Nevertheless, for the sake of completeness, Appendix A.3 illustrates the equilibrium dynamic with sectoral factor intensity differences.[18]

1.2.3 Equilibrium

Definition

In this economy, an equilibrium is defined as follows:

DEFINITION 1.1. *A dynamic competitive equilibrium is a time path of households' per-capita expenditure levels, wealth stocks and consumption quantities* $\left\{e_i(t), a_i(t), x_j^i(t)\right\}_{t=0}^{\infty}$, $j = 1, 2$, $\forall i$*; an evolution of prices, wage, interest and rental rate,* $\{P_j(t), w(t), r(t), R(t)\}_{t=0}^{\infty}$, $j = 1, 2$ *and a time path of factor allocations* $\{L_1(t), L_2(t), K_1(t), K_2(t), K_3(t)\}_{t=0}^{\infty}$*, which is consistent with household and firm optimization, perfect competition, resource constraints and market clearing conditions.*

In the following I illustrate the equilibrium as the outcome of decentralized markets. However, since all markets are complete and competitive the Welfare Theorems apply and the dynamic competitive equilibrium coincides with the solution to the social planner's problem.

Resource constraints and market clearing conditions

In equilibrium, capital and labor markets have to clear, i.e.

$$L(t) = L_1(t) + L_2(t), \text{ and } K(t) = K_1(t) + K_2(t) + K_3(t), \ \forall t. \qquad (1.20)$$

Market clearing in the good, service and investment good markets requires

$$Y_j(t) = X_j(t), \ j = 1, 2, 3, \ \forall t. \qquad (1.21)$$

Since the price of the investment good is chosen as a numéraire, asset market clearing implies

$$N(t)\int_0^1 a_i(t)di = K(t),\ \forall t. \tag{1.22}$$

Finally, the market rate of return of capital has to equalize the rental rate net of depreciations, i.e. $r(t) = R(t) - \delta$, $\forall t$.

Equilibrium dynamic

Under the choice of numéraire, perfect competition, resource constraints and the market clearing conditions, the equilibrium in production is characterized by the following lemma.

LEMMA 1.5. *Firm optimization implies at each date t,*

$$r(t) = A - \delta, \tag{1.23}$$

$$w(t) = A\frac{\alpha}{1-\alpha}\frac{K_1(t)+K_2(t)}{L(t)},\ j = 1,2, \tag{1.24}$$

$$P_j(t) = \exp\left[-g_j t\right]\left[\frac{A}{1-\alpha}\right]\left[\frac{K_1(t)+K_2(t)}{L(t)}\right]^{\alpha},\ j=1,2, \tag{1.25}$$

$$Y_j(t) = \exp\left[g_j t\right]\left[\frac{L(t)}{K_1(t)+K_2(t)}\right]^{\alpha} K_j(t),\ j=1,2, \tag{1.26}$$

and

$$\frac{K_1(t)}{L_1(t)} = \frac{K_2(t)}{L_2(t)} = \frac{K_1(t)+K_2(t)}{L(t)}. \tag{1.27}$$

Proof. Optimization implies that the marginal rate of technical substitution is equal to the relative factor price, i.e. $\frac{w(t)}{R(t)} = \frac{\alpha}{1-\alpha}\frac{K_j(t)}{L_j(t)}$, $j = 1,2$. With $R(t) = A$ and (1.20), this gives (1.23) and (1.27). Next, $R(t)$ has to equalize the valued marginal product across all sectors. This yields

$$R(t) = A = (1-\alpha)\left[\frac{L(t)}{K_1(t)+K_2(t)}\right]^{\alpha} P_j(t)\exp\left[g_j t\right],\ j=1,2,$$

where (1.27) has been used. Solving for $P_j(t)$ gives (1.25). Finally, with (1.27), the production functions can be rewritten as (1.26). Q.E.D.

The dynamic competitive equilibrium is fully characterized by equations (1.12)-(1.17) and (1.19)-(1.26). The endogenous variables are: $X_j(t)$ and $Y_j(t)$, $j = 1, 2, 3$; $a_i(t)$, $\forall i$; $E(t)$, $P_j(t)$, $j = 1, 2$; $w(t)$, $r(t)$, $L_j(t)$, $j = 1, 2$; $K(t)$ and $K_j(t)$, $j = 1, 2, 3$. $a_i(0)$, $\forall i$, is exogenously given.

When we solve for the dynamic competitive equilibrium, we obtain the following proposition.

PROPOSITION 1.2. *Suppose we have*

$$A - \delta - \rho + \epsilon g_2 > 0, \tag{1.28}$$

$$\rho > (1-\alpha)\epsilon\left[A-\delta\right] + n + \epsilon g_2, \tag{1.29}$$

$$\alpha^{\epsilon}\bar{l}^{\epsilon} \geq \frac{1-\epsilon}{1-\gamma}\beta\left[\frac{L(0)}{K(0)}\frac{A\left(1-(1-\alpha)\epsilon\right)}{\rho - n - \epsilon g_2 - \epsilon(1-\alpha)\left(A-\delta-n\right)}\right]^{\epsilon(1-\alpha)}, \tag{1.30}$$

and

$$\gamma\left[g_2 - g_1\right] - \epsilon\left[\frac{g_2 + (1-\alpha)\left[A-\delta-\rho\right]}{1-(1-\alpha)\epsilon}\right] \leq 0. \tag{1.31}$$

Then, there exists a unique dynamic competitive equilibrium path along which

(i) per-capita consumption expenditures, wages, aggregate capital and capital allocated to the consumption sectors grow at constant rates

$$g_E^* - n = g_w^* = \frac{A-\delta-\rho+\epsilon g_2}{1-(1-\alpha)\epsilon} > 0, \tag{1.32}$$

$$g_K^* = g_{K_1+K_2}^* = g_E^*. \tag{1.33}$$

The saving rate is constant and the real, investment good denominated interest rate is given by $A-\delta$. The prices of goods and services change at constant rates

$$g_{P_j}^* = -g_j + \alpha\left[g_E^* - n\right],\ j = 1, 2. \tag{1.34}$$

(ii) the expenditure share devoted to goods changes at constant rate

$$g_{S_1}^* = -\gamma \left[g_1 - g_2\right] - \epsilon \left[g_2 + (1-\alpha)\left[g_E^* - n\right]\right] \leq 0. \tag{1.35}$$

Capital and labor allocated to the goods sector grow at constant rates

$$g_{K_1}^* = g_K^* + g_{S_1}^* \leq g_K^* \leq g_{K_2}^*(t), \text{ and } g_{L_1}^* = n + g_{S_1}^* \leq n \leq g_{L_2}^*(t), \ \forall t. \tag{1.36}$$

The relative price between consumption goods and services changes at constant rate

$$g_{P_1}^* - g_{P_2}^* = g_2 - g_1. \tag{1.37}$$

Proof. See Appendix A.1.2. Q.E.D.

Proposition 1.2 demonstrates that the model reconciles structural change and changing relative prices at a sectoral level with balanced growth on the aggregate. Let us first focus on part (i) which illustrates that the model features on the aggregate the standard properties of neoclassical growth theory.

The per-capita growth rate is increasing in the marginal product of capital, A, and decreasing in the rate of time preference, ρ, and the depreciation rate, δ. Furthermore, the Kaldor facts hold. Total labor income, $w(t)L(t)$, and the total capital income net of depreciation, $rK(t)$, grow at the same constant rate g_E^* as aggregate output. Thus, the per-capita output growth rate, the capital-output ratio, the saving rate and the labor income share are constant. Moreover, the real, investment good denominated interest rate is equal to $A-\delta$. Since both consumption sector prices change at constant rates (see (1.34)), any price index with constant sectoral weights grows at a constant rate too. Hence, deflated by any price index with constant weights, the real per-capita expenditure growth rate and real interest rate would be constant. In an economy with structural change, however, the sectoral weights of the true cost of living price index adjust over time. This would yield a non-constant growth rate of the true cost of living price index. But typically, changes in the growth rate of the price index due to weight adjustments are very small (see Ngai and Pissarides, 2004).[19]

The model exhibits no transitional dynamic and can be solved analytically.[20] Without exogenous TFP growth (i.e. with $g_1 = g_2 = 0$), the aggregate behavior would be the same as in Rebelo (1991). However, the intertemporal substitution elasticity of expenditure, $\frac{1}{1-\epsilon}$, is tied together with the expenditure elasticity of demand for goods, ϵ.[21]

Noteworthy, although preferences are non-Gorman and inequality matters, the Kaldor facts hold irrespective of the distribution of the expenditure levels. This holds true since the marginal propensity to save out of capital income is the same at all wealth levels (and the marginal propensity to save out of labor income is zero for all households). An unforeseen shock on the wealth distribution would change the demand structure, but not the aggregate saving rate. Consequently, capital accumulation, growth and the pace of structural change would be unaffected.

Part (ii) of Proposition 1.2 emphasizes the equilibrium's non-balanced features on the sectoral level. Although the Kaldor facts hold, the aggregate expenditure share devoted to goods as well as the relative price between goods and services change over time. The functional forms the simple model imposes are notable too. The model predicts that both the expenditure share of goods and the relative price of goods decrease at constant rates. Remarkably, this is consistent with the functional form of the stylized facts depicted in Figure 1.1 and 1.2.

The shift in the aggregate demand structure transmits to the production side (see (1.36)). Capital allocated to the goods sector grows at a lower rate than the aggregate capital stock, which itself grows at a lower rate than capital allocated to the service sector. In contrast to $g^*_{K_1}$ and g^*_K, $g^*_{K_2}(t)$ expands at a time varying rate. The same applies to the allocation of labor. If n is small relative to $g^*_{S_1}$, the absolute quantity of labor allocated to the goods sector can even decrease. Nevertheless, consumption of both goods and services increase steadily – even in per-capita terms. Thus, the goods sector declines only in relative and not in absolute terms.

The required parametric restrictions (1.28)-(1.31) are harmless. Reconciliation of the non-balanced features of growth with the Kaldor facts does not depend on any knife-edge condition. (1.28) ensures positive capital accumulation and growth in per-capita terms. Condition (1.29) is necessary and sufficient for the transversality condition to hold. Furthermore, it is also sufficient to ensure finite utility. Condition (1.30) makes sure that condition (1.3) is met for all households at $t = 0$. Moreover, together with condition (1.31), it ensures condition (1.3) along the entire equilibrium path.

In general, the structural change is driven by income and substitution effects. With $\epsilon > 0$ services are luxuries. Hence, due to per-capita growth, the expenditure share devoted to services tends to increase. In addition, if the relative price changes (i.e. $g_1 \neq g_2$), there is a substitution effect. Since the elasticity of substitution between the two consumption sectors is strictly less than one, the expenditure share of the sector with the higher TFP growth rate tends to decrease. The magnitude of the income and substitution effects is controlled by the exogenous preference parameters γ and ϵ. With $\epsilon = 0$ we have homothetic preferences and changes in expenditure shares are exclusively determined by the substitution effect. With $g_1 = g_2$ the relative price does not change and the entire structural change is driven by an income effect. In general, income and relative price effects can go in opposite directions. If, by sheer coincidence $-\gamma(g_1 - g_2) = \epsilon\,[g_2 + (1-\alpha)\,[g_E^* - n]]$, the two effects cancel each other so that there would be no structural change.[22]

In the next proposition the income and substitution components of structural change and the model's cross-sectional predictions are analyzed in more detail.

PROPOSITION 1.3. *Along the equilibrium path,*

(i) for all households, the expenditure share devoted to goods changes at a constant rate $g_{S_1}^* \leq 0$.

(ii) *according to the substitution effect, a decrease of the relative price of goods by one percent, decreases the expenditure share devoted to goods of household i by $-\gamma + \epsilon s_1^i(t) \leq 0$ percents.*

(iii) *for all households, according to the income effect, an increase of the per-capita expenditure level by one percent, decreases the expenditure share devoted to goods by ϵ percents.*

Proof. Part (i) follows from (1.6) and the fact that $g_{e_i} = g_E^* - n$, $\forall i, t$. $s_1^i(t)$ can be written in terms of prices and attained utility level, $V_i(t)$, as (see (A.1) and (1.6))

$$s_1^i(t) = \beta \left[\epsilon \left[V_i(t) + \frac{\beta}{\gamma} \left[\frac{P_1(t)}{P_2(t)} \right]^\gamma + \frac{1}{\epsilon} - \frac{\beta}{\gamma} \right] \right]^{-1} \left[\frac{P_1(t)}{P_2(t)} \right]^\gamma .$$

For the elasticity of $s_1^i(t)$ with respect to $\frac{P_1(t)}{P_2(t)}$ we then get $-\gamma + \epsilon\beta \left[\frac{P_2(t)}{e_i(t)} \right]^\epsilon \left[\frac{P_1(t)}{P_2(t)} \right]^\gamma$, or $-\gamma + \epsilon s_1^i(t)$, which is non-positive since $s_1^i(t) \leq 1$ and $\gamma \geq \epsilon$. Part (iii) follows immediately from (1.6). Q.E.D.

The model predicts that not only the aggregate, but also all individual expenditure shares of goods decrease at the identical, constant rate $g_{S_1}^*$. This is consistent with the linear and parallel decline of the logarithmized expenditure shares of different income quintiles (see Figure 1.3). However, as part (i) and (ii) of Proposition 1.3 show, if $\epsilon > 0$, the division of this change in expenditure shares into an income and substitution effect differs across households. For richer households (with a lower $s_1^i(t)$), the substitution effect is relatively more important. Consequently, as all $s_1^i(t)$ decline, the relative importance of the income effect as a determinant of the aggregate structural change decreases over time. Since preferences allow for a representative agent in Muellbauer's sense, the substitution effect of the aggregate economy is the same as the substitution effect for the representative agent. Hence, a one percent decline in the relative price of goods decreases (according to the substitution effect) the aggregate expenditure share of goods by $-\gamma + \epsilon S_1(t) \leq 0$ percents.

An alternative way to illustrate how well the model fits the cross-sectional data is to look at the suggested relationship between the expenditure structure and the per-capita expenditure level. Logarithmizing both sides of (1.6) gives

$$\log s_1^i(t) = b(t) - \epsilon \log e_i(t), \tag{1.38}$$

where $b(t) \equiv \log [\beta P_2(t)^{\epsilon-\gamma} P_1(t)^{\gamma}]$. Consequently, the model predicts – after allowing for a time dependent intercept $b(t)$ – an iso-elastic relation between the expenditure share of goods and the per-capita expenditure level of different households. Figure 1.6 depicts the partial correlation between the logarithm of these two variables for the income quintiles already considered in Figure 1.3. It is striking how well a linear line approximates the relationship.

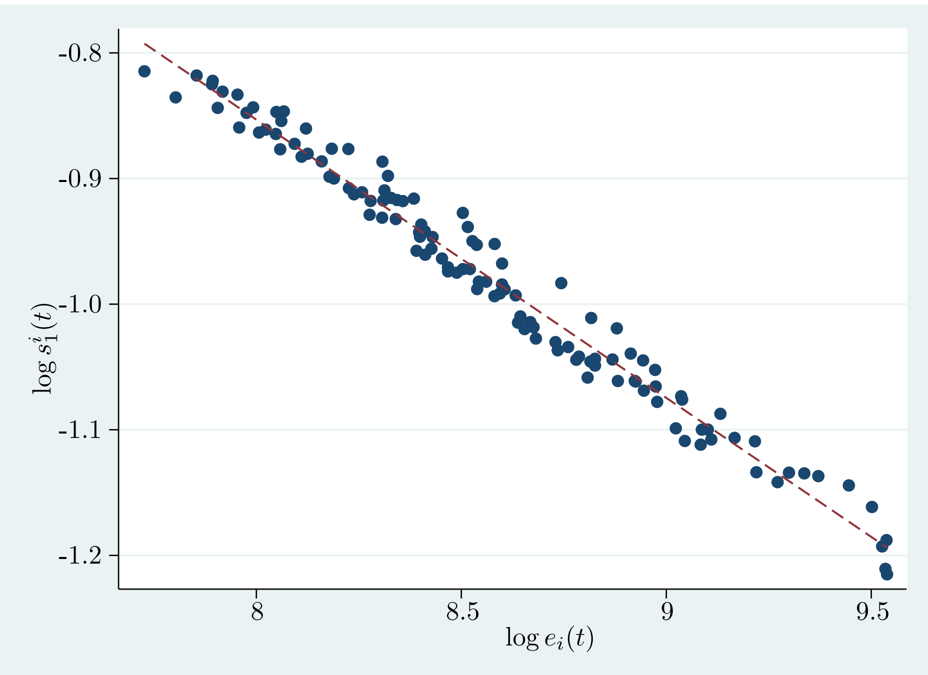


Figure 1.6: Scatter plot of cross-sectional variation

Notes: The figure depicts the partial correlation between the logarithmized expenditure level per-equivalent scale and the logarithmized expenditure share of goods of a given income quintile, where we allowed in each year for a separate (distinct) intercept. The slope of the fitted line is -0.2214. This slope is the same as if we regress the logarithmized expenditure share on the logarithmized expenditure level per equivalent scale and time dummies. The R^2 of this underlying regression is 0.9494 and the standard error of the slope coefficient is 0.0042. Source: Consumer Expenditure Survey.

It is insightful to take a closer look at the equilibrium toward which the economy converges, as time goes to infinity. To do so, we define:

DEFINITION 1.2. *The asymptotic equilibrium is the dynamic competitive equilibrium path toward which the economy tends as time goes to infinity.*

Then, we have the following proposition (asymptotic equilibrium values are denoted by a superscript A).

PROPOSITION 1.4. *Suppose, condition* (1.31) *holds with strict inequality (i.e. there is structural change). Then, in the asymptotic equilibrium,*

(i) the expenditure share devoted to goods is equal to zero, i.e. $S_1^A = 0$.

(ii) the expenditure elasticity of demand is $1 - \epsilon$ *for goods and unity for services.*

(iii) the elasticity of substitution between goods and services, σ_i^A*, is equal to* $1 - \gamma$ *for all households* i.

Proof. Since (1.31) holds with strict inequality S_1 converges to 0 (see (1.35)) and the elasticities of Lemma 1.3 converge to the corresponding values. Q.E.D.

Part (i) of Proposition 1.4 shows that the service sector is the asymptotically dominant consumption sector. The existence of an asymptotically dominant sector is a common feature of the models by Ngai and Pissarides (2007), Acemoglu and Guerrieri (2008) and Foellmi and Zweimueller (2008). The asymptotic dominance of the service sector is not a fact of a trivial disappearance of the good sector. In absolute terms, the asymptotically consumed quantity of goods goes to infinity – even in per-capita terms.

Part (ii) and (iii) of Proposition 1.4 illustrate how parsimonious the model is. The expenditure elasticity of demand and the elasticity of substitution across sectors control size and magnitude of relative price and income effects on S_1. The model has exactly two exogenous parameters, ϵ and γ, which control separately the asymptotic values of these two elasticities. In general, with $\epsilon \neq 0$ and $g_1 \neq g_2$, both income and relative price effects are even asymptotically present (note that all the properties stated in Proposition 1.2 hold asymptotically too). With $\epsilon = 0$ the asymptotic equilibrium is similar to the one by Ngai and Pissarides (2007) and Acemoglu and Guerrieri (2008). There is no income effect and the elasticity of substitution across sectors is constant. With $g_1 = g_2$, there is no relative price effect and the asymptotic equilibrium resembles the one by Foellmi and Zweimueller (2008). But in contrast to Foellmi and Zweimueller (2008), where the expenditure elasticity of demand of the asymptotically dominated sectors converge to zero, it can in this model be set to any value between 0 and 1.[23]

So far, it has been shown that the model is consistent with a unique dynamic competitive equilibrium path, along which the Kaldor facts hold and changes in expenditure shares and relative prices occur. Furthermore, the functional form of these nonbalanced features is consistent with the dynamics observed in the U.S. data on the aggregate as well as on the cross-sectional level. Two model parameters – ϵ and γ – determine the magnitude of the income and substitution effect on the structural change. It is the aim of the next section to quantify these two forces.

1.3 Empirical quantification

1.3.1 Quantitative replication of the structural change

According to the theoretical model of section 1.2, the structural change in aggregate expenditures is described by (see (1.15))

$$g^*_{S_1} = -\epsilon \left(g^*_E - g^*_{P_2} - n \right) + \gamma \left(g^*_{P_1} - g^*_{P_2} \right). \tag{1.39}$$

In this expression we already made use of the constancy of the involved growth rates, which is the model's general equilibrium implication (see Proposition 1.2). The data suggests that the growth rate of the expenditure share devoted to goods, $g^*_{S_1}$, is -0.010, the growth rate of per-capita expenditures in terms of services, $g^*_E - g^*_{P_2} - n$, is 0.016 and the growth rate of the price of goods relative to services, $g^*_{P_1} - g^*_{P_2}$, is -0.016.[24] When we plug these values into (1.39), we conclude that the model is quantitatively consistent with the observed structural change, growth and relative prices as long as the (ϵ, γ)-combination fulfills

$$\epsilon + \gamma = 0.625. \tag{1.40}$$

1.3.2 Estimating ϵ and γ

Equation (1.40) is uninformative about the relative importance of the substitution and income effects. However, with equation (1.38) the theoretical model makes a very precise prediction about the cross-sectional variation in the expenditure structure. In order to identify ϵ, this suggests to regress the logarithmized expenditure share of goods on a time fixed effect and the logarithmized expenditure level. But there arises one additional difficulty with this regression. Expenditures classified as "goods" include some quantitative important durable items as cars or furniture. And we observe in the Consumer Expenditure Survey a household's expenditures for only a relatively short period of time (up to a maximum spell of 4 quarters). Hence, in the simple regression, households which happen to buy a new car in the observed quarter have very high per-capita expenditures and would (wrongly) be considered as extraordinary rich. Since buyers of a new car have at the same time an exceptionally high goods share, the simple estimate for ϵ is biased towards zero. As a solution, I use the logarithm of the household's yearly after tax labor income plus transfers per equivalent scale as an instrument for the logarithmized per-capita expenditure level.[25] The results obtained by this IV approach are summarized in Table 1.1. The estimate for ϵ is always positive and statistically highly significant. When we additionally control

Table 1.1: Cross-sectional estimation of ϵ

Dependent variable: $\log s_1^i(t)$				
	(1)	(2)	(3)	(4)
$-\log e_i(t)$	0.181***	0.205***	0.218***	0.230***
	(0.002)	(0.002)	(0.002)	(0.002)
Children share		0.203***	0.125***	0.125***
		(0.003)	(0.004)	(0.005)
Elderly share		−0.077***	−0.083***	−0.055***
		(0.003)	(0.003)	(0.004)
Residence indicators	No	No	Yes	Yes
Family size indicators	No	No	Yes	Yes
Ref. person controls	No	No	No	Yes
Observations	450,602	450,602	404,079	404,079
R^2	0.012	0.026	0.031	0.036
Method	IV	IV	IV	IV

Notes: Standard errors in parentheses. *** significant at 1 percent, ** significant at 5 percent, * significant at 10 percent. All regressions include quarter fixed effects (96 groups). The logarithmized expenditure level per equivalent scale is instrumented by the logarithmized after tax labor earnings plus transfers per equivalent scale. “Children share” and “Elderly share” measures the share of household members with age < 18 and ≥ 65, respectively. “Residence indicators” consists of regional dummies (4 groups), a rural/urban dummy as well as indicators of different population density of the city of residence (5 groups). “Family size indicators” consists of 11 groups. “Ref. person controls” consists of the age, the sex and a race indicator (4 groups) of the reference person.

for other household and reference person characteristics the estimate for ϵ increases slightly above 0.2 (see column (2) to (4)).[26]

Hence we conclude that the cross-sectional data allows us to identify ϵ and suggests that a value of about 0.22 is reasonable. This value implies an expenditure elasticity of demand for goods of 0.78. An alternative way to infer who reasonable this parameter value is, is via the implied elasticity of substitution. With $\epsilon = 0.22$, a replication of the structural change implies for γ a value of 0.405 (see (1.40)). According to Proposition 1.4, $1 - \gamma$ can be interpreted as the asymptotic value of the elasticity of substitution. Hence, with $\gamma = 0.405$ the elasticity of substitution of the representative agent converges (from below) to 0.596. This value is in the range of other estimates and calibrations of the elasticity of substitution (see footnote 9).[27]

This highlights that both channels of structural change are of empirical importance. The model could potentially generate the observed structural change with an income effect alone (and an asymptotic elasticity of substitution equal to unity). But this would require an ϵ of 0.625 (see 1.40), denoting an expenditure elasticity of demand for goods of $1-\epsilon = 0.375$. Such a strong income effect is clearly at odds with the cross-sectional data. Conversely, however the homothetic case with $\epsilon = 0$ is also clearly rejected by the data. With the parameter values $\epsilon = 0.22$ and $\gamma = 0.405$ the model suggests that in 1946, 44 percent of the observed structural change is attributed to a relative price effect, whereas the remaining 56 percent are attributed to the income effect.[28] In 2011, the corresponding numbers are 53 percent and 47 percent, respectively. Furthermore, the model predicts that the relative contribution of the substitution effect will asymptotically converge to 65 percent.[29]

1.4 Conclusion

This paper presented a parsimonious growth theory, which is consistent with structural change, relative price dynamics and the Kaldor facts. The model allows us to analyze both explanations of structural change – income and substitution effects – simultaneously. To the best of my knowledge, such a theory did not exist yet.

The virtues of the theory are twofold. First, the model's functional form fits the data very well and the framework can replicate the observed structural change quantitatively. Moreover, not only the model's predicted dynamic of the aggregate expenditure shares, but also the predicted cross-sectional variation is confirmed by the data. And the paper shows how this cross-section variation can be exploited to estimate the model's key parameters and quantify the two driving forces of structural change.

The second virtue is given by the exact replication of the Kaldor facts, which is clearly desirable from an empirical point of view. In the data we see a fast and persistent structural change. Reconciling this with a relatively stable interest, saving and aggregate growth rate is challenging. Although some calibrations of models of structural change are approximately consistent with the Kaldor facts, others are clearly not. This paper suggests that this shortcoming is mainly an artifact of the functional form of the specified intratemporal utility function.

Additionally, the exact replication of the Kaldor facts is very appealing from a theoretical perspective too. Structural change is interrelated to many important aspects of demographics, labor supply, income inequality and convergence, international trade or biased technical change. These phenomena are often outlined in standard one-sector neoclassical growth models (with balanced growth). To analyze them in a multi-sector model, a tractable theory of structural change is just a starting point. I hope the presented framework provides to be useful in order to study these important questions.

Notes

1. Changes in relative prices affect the expenditure structure whenever the elasticity of substitution across sectors is unequal to unity. This mechanism of structural change goes back to Baumol (1967), who emphasizes total factor productivity (TFP) growth differences as a source of relative price changes. In Acemoglu and Guerrieri (2008), capital deepening and sectoral factor intensity differences is another determinant of the relative price dynamic. But in contrast to Ngai and Pissarides (2007) the Kaldor facts hold only asymptotically.

2. Note that a constant elasticity of substitution implies that relative nominal expenditures are an iso-elastic function of the relative price, where the elasticity is one minus the elasticity of substitution.

3. Acemoglu and Guerrieri (2008) conclude: "It would be particularly useful to combine the mechanism proposed in this paper with nonhomothetic preferences and estimate a structural version of the model with multiple sectors using data from the U.S. or the OECD."(Acemoglu and Guerrieri, 2008, p. 493).

4. This mechanism of structural change is consistent with Engel's law, which is regarded as one of the most robust empirical regularity in economics (see Engel, 1857; Houthakker, 1957; Houthakker and Taylor, 1970 and Browning, 2008). As a consequence, many models of structural change rely on income effects. See e.g. Matsuyama (1992), Echevarria (1997), Laitner (2000), Caselli and Coleman (2001), Kongsamut, Rebelo and Xie (2001), Gollin, Parente and Rogerson (2002) and Greenwood and Seshadri (2002) which use quasi-homothetic intratemporal preferences or Falkinger (1990), Falkinger (1994), Zweimueller (2000), Matsuyama (2002), Foellmi and Zweimueller (2008) and Buera and Kaboski (2009b), which generate non-homotheticity by a hierarchy of needs.

5. In Kongsamut, Rebelo and Xie (2001) consistency with the Kaldor facts relies on a widely criticized knife-edge condition, which ties together preference and technology parameters and implies constant relative prices. Foellmi and Zweimueller (2008) have to assume that technological differences are uncorrelated with the hierarchical position of a good (and its sectoral classification).

6. See also the recent empirical works by Buera and Kaboski (2009a) and Herrendorf, Rogerson and Valentinyi (2009), which estimate the relative contribution of income and substitution effects for the U.S.

structural change. In contrast to these two papers, the structural estimation of this work is based on a preference specification which is consistent with the Kaldor facts. Moreover, its is an explicit ambition of this paper to be consistent with the cross-sectional (expenditure) data.

7. For $\epsilon = 0$ we get the limit case with $V(\cdot) = \log\left[\frac{e_i(t)}{P_2(t)}\right] - \frac{\beta}{\gamma}\left[\frac{P_1(t)}{P_2(t)}\right]^{\gamma} + \frac{\beta}{\gamma}$ and with $\gamma = \epsilon = 0$ we would obtain Cobb-Douglas preferences with $V(\cdot) = \log\left[\frac{e_i(t)}{P_1(t)^{\beta}P_2(t)^{1-\beta}}\right]$. As another special case, with $\beta = 0$, we would have only one consumption sector and CRRA preferences.

8. Appendix A.2 shows that the class of preferences specified in this paper is the most general class of intratemporal preferences defined over two sectors implying a behavior which is jointly consistent with a constant (negative) growth rate of the expenditure share devoted to one sector (see Figure 1.1) and a constant (positive) growth rate of per-capita expenditures (one of the Kaldor facts) in an environment where the relative price changes at a constant rate too (see Figure 1.2).

9. Acemoglu and Guerrieri (2008) and Buera and Kaboski (2009a) calibrate their models with an elasticity of substitution equal to 0.76 and asymptotic 0.5, respectively. And in Herrendorf, Rogerson and Valentinyi (2009) the model's best fit of final consumption shares is attained with an asymptotic elasticity of substitution equal to 0.81 (or 0.52, respectively if government consumption is excluded). Furthermore, the elasticity of substitution between goods and services has been of interest in international macroeconomics in order to use it as a proxy for the elasticity of substitution between tradable and non-tradable commodities. Also in this literature the elasticity of substitution has consistently been estimated to be lower than unity (see e.g. Stockman and Tesar, 1995 who obtain a value of 0.44).

10. The utility function (1.2) could also generate cases where the expenditure elasticity of demand for goods or the elasticity of substitution exceeds unity. But because they are not empirically relevant, these cases where excluded by the restriction $0 \leq \epsilon \leq \gamma < 1$.

11. For $\epsilon = 0$, we have - according to Muellbauer's definition - the limit case with $e^{RA}(t) = \frac{E(t)}{N(t)}$.

12. With $\epsilon > 0$, a high dispersion of per-capita expenditure levels is associated with a low value of ϕ. In the homothetic case, we have a representative agent economy (in the narrower sense), where inequality does not matter (i.e. $\phi = 1$).

13. In contrast to models with 0/1 preferences and intertemporal optimization (see e.g. Foellmi and Zweimueller, 2006 and Foellmi, Wuergler and Zweimueller, 2009) this model focuses on the intensive margin of consumption. Moreover, the model here allows us to study any – possibly continuous – income distribution with a lower bound such that condition (1.3) is fulfilled.

14. Appendix A.4 shows how these sector specific TFP growth rates can be endogenized.

15. With $\beta = 0$ and $g_2 = 0$ the model would coincide with the one by Rebelo (1991).

16. Where – as specified above $F(K_j(t), \exp[g_3 t] L_j(t))$ is the neoclassical production function of the investment sector. This approach is chosen by Kongsamut, Rebelo and Xie (2001) and by Foellmi and Zweimueller (2008). In addition, they both (have to) assume that $A_j(t)$, $j = 1, 2$ is constant over time.

17. In contrast to this, $A_1(t)$ could follow any process and aggregate growth would still be balanced. But in order to be consistent with the data presented in Figure 1.1 and 1.2 (and also in line with the large body of the literature) productivity growth is assumed to occur in the good sector at a constant rate too.

18. In this case the model is relatively similar to the one by Acemoglu and Guerrieri (2008) and the Kaldor facts hold only asymptotically. However, in this paper, structural change is also determined by an income effect.

19. The growth rate of the partial true cost of living price index of household i is defined as $g_P^{TCL}(t) = g_{P_2}(t) + s_1^i(t)\left[g_{P_1}(t) - g_{P_2}(t)\right]$ (see Pollak, 1975). In the data, relative price growth rate is -1.6 percent

and in 2011 the aggregate expenditure share of goods was 0.34, whereas its asymptotic value is zero. Hence, measured by the true cost of living price index of the representative household, the model predicts the real interest rate in 2011 to be 0.005 higher than its asymptotic value.

20. This is due to the AK specification of the production function of investment goods. With a decreasing marginal product of capital, transitional dynamics would arise.

21. With $\epsilon = 0$, this interdependence reflects the result obtained by Ngai and Pissarides (2007): If preferences are homothetic, reconciliation of structural change with the Kaldor facts requires that the intertemporal substitution elasticity of expenditures is equal to unity.

22. A trivial case, where this condition is fulfilled arises if neither an income nor a substitution effect exists. This occurs with homothetic preferences and a constant relative price ($\epsilon = g_1 - g_2 = 0$) or with Cobb-Douglas preferences ($\epsilon = \gamma = 0$).

23. This flexibility is also an important difference to theories relying on generalized Stone-Geary preferences, where the asymptotic expenditure elasticity of demand is unity for all sectors. This asymptotic inexistence of income effects leads to a suboptimal fit of the data, as Buera and Kaboski (2009a) show in their calibration: "The model fails to match the sharper increase in services and decline in manufacturing after 1960. [...] Explaining this would require a large, delayed income effect toward services. This is not possible with the Stone-Geary preferences, where the endowments and subsistence requirements are most important at low levels of income." (Buera and Kaboski, 2009a, p. 473-474.)

24. See Figure 1.1 and 1.2 as well as Figure A.1 in Appendix A.5, which also illustrate how well the constant growth rates approximate the three series.

25. This solves the problem since in quarters in which households buy a new car the labor income is – in contrast to total expenditures – not (by construction) above its average. An alternative approach would be to group households according to their income. As it can be inferred from Figure 1.6 or table 1 in the earlier version of this paper (see Boppart, 2011) this leads us to very similar estimates for ϵ. An advantage of the IV regression is that it allows us to control for additional individual household characteristics.

26. Figure A.2 in Appendix A.5 shows the estimates for ϵ if we run the regression of column (4) in Table 1.1 for each year separately. $\hat{\epsilon}$ is very stable over time and apart from two exceptions always between 0.20 and 0.25.

27. Moreover, the combination $\epsilon = 0.22$ and $\gamma = 0.405$ fulfills the assumed parametric restriction $0 \leq \epsilon \leq \gamma < 1$.

28. In 1946, the goods sector accounted for 60 percent of total personal consumption expenditures. Then, the change in expenditure share attributed to the substitution effect is equal to an annualized rate of $(-0.405 + 0.22 \cdot 0.6) \cdot 1.6 = -0.435$ (see Proposition 1.3).

29. For this numerical exercise we just had to pin down the two preference parameters ϵ and γ. A full calibration of the model is provided in the appendix of Boppart (2011).

Chapter 2

Online accessibility of academic articles and the diversity of economics

Chapter Summary

A key aspect of generating new ideas is drawing from different elements of preexisting knowledge and combining them into a new idea. In such a process, the diversity of ideas plays a central role. This paper examines the empirical question of how the internet affected the diversity of new research by making the existing literature accessible online. The internet marks a technological shock which affects how academic scientists search for and browse through published documents. Using article-level data from economics journals for the period 1991 to 2009, we document how online accessibility lead academic economists to draw from a more diverse set of literature, and to write articles which incorporated more diverse contents.

2.1 Introduction

Two elements broadly characterize academic research: (*i*) the production of knowledge in a "recombinant growth" framework (Weitzmann 1996, 1998a) where new ideas represent innovative combinations of previous ones, and (*ii*) attention as a scarce resource which limits researchers' processing capacity of existing ideas (Franck, 1999, Klamer and Van Dalen, 2002, Falkinger, 2007b).

Recombinant growth stresses the notion that the heterogeneity of existing ideas which serve as input is positively linked to knowledge accumulation.[30] The French mathematician Henri Poincaré described this idea succinctly as (Poincaré, 1910, p. 325): "Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart." In such a process the preservation of diversity of academic publications is important because a more diverse stock of knowledge enhances research productivity. Limited attention, on the other hand, implies that what matters specifically in this context is the diversity *perceived* by researchers. As Weitzman (1998a, p. 333) puts it: "[T]he ultimate limits to growth may lie not so much in our ability to generate new ideas, so much as in our ability to process an abundance of potentially new seed ideas into usable form."[31] This local diversity (i.e. diversity perceived by an individual researcher) depends on characteristics of researchers and on features of the technology used by researchers to learn about existing ideas.

In this paper, we explore the effect of one such technological aspect, the digitization of academic literature and its dissemination through the internet. The internet marks a profound technological shock which affected the way academic scientists search for and browse through published documents. On the one hand, the internet exemplifies a huge increase in the availability of scientific articles at very low (time) cost. On the other hand, the internet offers very powerful new tools such as search engines and hyperlinks to browse through this sheer amount of information. The specific empirical question we ask is how the internet affected the diversity of newly undertaken research by making the existing literature accessible online. The answer to this question is far from obvious. The new tools and, especially, search algorithms may allow researchers to find forgotten “lost pearls” and bring to their attention contemporaneous articles they would not read habitually. However, the internet also entails –almost by an empirical law– very unequal distributions of attention (Huberman, 2003). Evans (2008) documents that as more scientific articles became available online more recent articles were referenced more often and citations were more concentrated on fewer documents. In contrast to these aggregate measures of diversity, we focus on local ones measuring how diverse the ideas are a publication contains or is based on. Our results show that these local measures of diversity increase with the share of relevant literature being accessible online. These empirical findings can be linked to theoretical models of attention economies where comparative statics predict that increases in the diffusion of sender signals may diminish the equilibrium number of senders while resulting in access to a larger variety of senders for each individual receiver (Falkinger, 2007a, 2008).

The data we analyze consists of roughly two decades of publications in core economics journals, starting in the pre-internet era and including the complete transition to full digitization. Our paper exploits the same basic exogenous variation in the date of online publication across different journals and across volumes within journals pioneered by Evans (2008), and contributes to a small and very recent literature relying on the same source of variation which explores the impact of online access for the economics profession (Depken and Ward, 2009; McCabe and Snyder, 2011).[32] So far, this literature has focused primarily on the impact that articles’ online access had on these articles’ number of citations. Depken and Ward (2009) show that access to the online platform “Journal STORage” (JSTOR) increases the number of citations to journals contained in JSTOR as well as to older journal volumes. McCabe and Snyder (2011) document that the number of citations a publication receives increases by about 10 percent as it is included in JSTOR and that this effect is about the same both for often-cited as well as rarely-cited papers. By studying the impact on upcoming articles' contents, our research addresses quite a different aspect of the scientific process.

Section 2.2 introduces our two measures of diversity. The construction of our measure of online accessibility and our identification strategy are outlined in detail in section 2.3. Estimation results are discussed in sections 2.4 and 2.5 and section 2.6 concludes.

2.2 Two measures of diversity

Our analysis considers two measures of diversity: (*i*) the distribution of pairwise geodesic distances of cited references and (*ii*) the number of Journal of Economic Literature (JEL) classification codes assigned to a publication. The geodesic distance between two items is the shortest back-in-time connection within the citation network.

The distribution of geodesic distances between the references of an article is a distinct measurement of diversity: The share of short distances is high if a publication draws only from one narrow and well-connected literature. In contrast, higher shares of large distances result if the paper connects several separated strands of the literature for the first time. The second measure of diversity considered is the JEL codes assigned to a publication by the American Economic Association's bibliography, EconLit.[33] The JEL classification indexes the contents of an article describing which fields and subfields it falls into, and thus uncovers the breadth of an article within economics. In the following, these two measures are explained in more detail.

The measures were obtained for every article published between 1991 and 2009 in 50 selected core journals of economic research. The list of journals includes all "top five,"[34] top field and second tier general interest journals (as well as their historical predecessors). Table B.9 in Appendix B.1 provides an alphabetic list of the journals.[35]

2.2.1 Geodesic distances

Geodesic distances provide essential information about the structure of networks. Measures of network connectivity and centrality are usually characterized by functions of these distances.[36] The analysis of geodesic distances in networks of academic citations dates back at least to De Solla Price (1965). The previous literature is largely explorative and aims at describing the distribution of geodesic distances in particular citation networks (e.g. Yin et al., 2006, Franceschet, 2012).[37,38] By focusing on the distribution of geodesic distances of *articles' references*, we capture the local connectivity of references, which we interpret as a measure of articles' (local) diversity. This local connectivity specific to citation networks has not been explored in the literature so far.

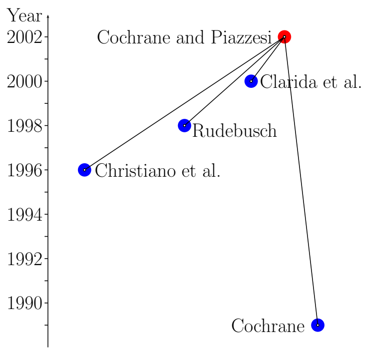
The calculation of geodesic distances requires knowledge of the entire network of citations. We downloaded from Thomson Reuters' Web of Science the list of references of all items published between 1955 and 2009 in the 50 core economics journals. The sample does not only include articles but also notes, letters, book reviews etc., which gives rise to a total of 129,145 items. To construct the citation network, references were matched back to the published items. On average, we are able to match 36 percent of all references, and 44 percent of references in articles published 1991 to 2009. Unmatched references may refer to publications prior to 1955 or to publications in books, working papers or disregarded journals. Finally, we calculated the shortest back-in-time connection within the citation network for all binary pairs of (identified) references.

In Figure 2.1, this process is illustrated for a paper-and-proceedings article written by John Cochrane and Monica Piazzesi and published in 2002,[39] visualized in the graphic in Panel (a) by a red node. This article made seven references. Within our sample we can identify four of them and these items are depicted as blue nodes.[40] The four identified references give rise to six bilateral connections (unordered pairs) among them. We then calculated for each of the bilateral links the shortest back-in-time connection within the entire citation network spanned by the sample of over 100,000 items. In the example of Figure 2.1, one of the references, *Rudebusch (1998)*, cites another one directly –*Christiano et al. (1996)*– which implies a geodesic distance of one (Panel b). Moreover, *Christiano et al. (1996)* is linked to a third reference, *Cochrane (1989)*, via two connections; as is the case for *Rudebusch (1998)* and the fourth reference, *Clarida et al. (2000)*. Panel (c) plots these geodesic distances of order two. The shortest connection from *Clarida et al. (2000)* to *Cochrane (1989)*, and to *Christiano et al. (1996)*; as well as between *Rudebusch (1998)* and *Cochrane (1989)*, is given by three steps (Panel d). We determine the geodesic distances iteratively up to a length of 3, giving rise to a probability mass function over four categories, with the last category comprising geodesic distances strictly larger than three. In the example of Figure 2.1, the fraction of pairwise geodesic distances of order one, two, three and higher than three are $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, and 0, respectively.

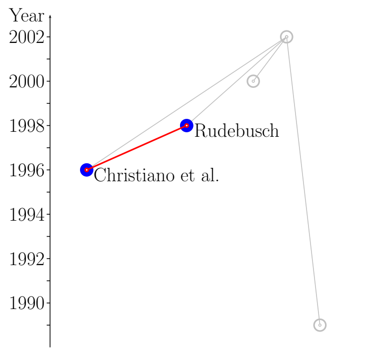
This simple example, chosen for illustration purposes, is not typical for the dataset. The median article has 10 identified references and thus its references' citation network comprises 45 pairs. Over the whole sample period 1991-2009, the average fractions of these geodesic distances are 25, 27, 20, and 28 percent.

2.2.2 JEL codes

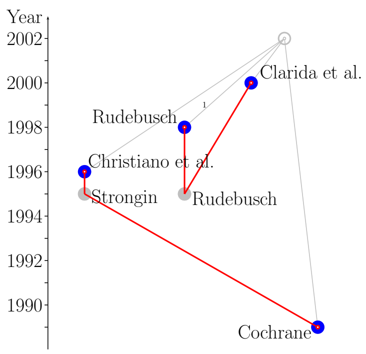
Our second measurement of diversity is the number the JEL classification codes assigned to a publication. Up to six three-digit JEL codes are assigned by EconLit to each article and we downloaded this information from the EconLit webpage. The first digit of a JEL codes is a letter which divides economics into twenty main fields, such as "public economics" or "industrial organization". The JEL classification system was introduced in 1991 and consequently we observe these classification codes only from then onwards.[41] While half the articles fall into exactly one field according to the one-digit definition, about 37 percent contribute to two fields, and somewhat over 10 percent have three one-digit JEL codes. There are large differences between



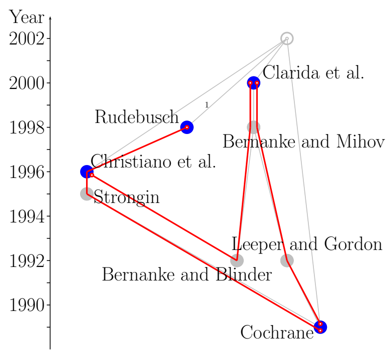
(a) Article with references



(b) Geodesic distance of order one



(c) Geodesic distances of order two



(d) Geodesic distances of order three

Figure 2.1: Geodesic distances of an article's references

Notes: The Figure illustrates how geodesic distances of an article's references are obtained, using the article by John Cochrane and Monica Piazzesi, "The Fed and Interest Rates: A High-Frequency Identification", *American Economic Review, Papers and Proceedings*, May 2002, Volume 92, Issue 2, pp. 90-95. Panel (a) plots the article as a red node and the four references identified in the data as blue nodes. Panels (b), (c) and (d) plot shortest back-in-time citation paths between blue nodes (geodesic distances) as red lines. Blue nodes' references relevant for these paths are plotted in grey.

journals; for instance, while the average article in Econometrica has about 1.1 one-digit JEL codes, the Journal of Development Economics' average article has about 2.3. The variation within journals is even larger. Relying on journal-year fixed effects, this variation within a journal is the one exploited in the empirical section. The last two digits classify the twenty fields into narrower sub- and subsubfields resulting in a very subtle measure of diversity. The median article has two three-digit JEL codes, while about one third of articles have more than two. In our analysis we consider the number of distinct one-, two- and three-digit JEL codes each as a separate dependent variable.[42]

JEL codes constitute a unique and precise categorization of articles' contents beyond its main field, which other, similarly structured applications such as patent citations lack. In such datasets the intellectual content of a patent is limited to one "patent class" only. An example is the NBER patent citations dataset (cf. Hall, Jaffe and Trajtenberg, 2001). Trajtenberg, Henderson and Jaffe (1997) introduced a variable called "originality" which is (the negative of the) Herfindahl concentration index of different patent classes a specific patent cites. Clearly, our diversity measures are closely related to this concept of originality.

2.3 Online accessibility of economics journals

For the online accessibility of publications we use data from two sources: "Fulltext Sources Online" (FSO) and JSTOR. The FSO data contains, for each year 1998-2009 and each online platform, information on which volumes of which journal were accessible online.[43] The FSO data contains this information for the journals' own webpage as well as for all major providers (such as e.g. EBSCOhost, LexisNexis, ScienceDirect or WilsonWeb) with the important exception of JSTOR.[44] Since JSTOR is one of the most important providers of online access (and has been even more important historically) we augment the dataset with the information about the date of a journal volume's first download at JSTOR.

A satisfactory measure of online accessibility based on these data should enable us to distinguish online accessibility from other secular time trends such as general internet usage. Achieving this should be possible since research projects differ in which subsets of the entire past literature are relevant to them. We assume that articles cite all relevant past works (a requirement stated in all journals' article submission guidelines) and define an article's relevant (past) literature as the set of journals in which the article's references were published.[45] Journals varied widely and unsystematically regarding the date when they first went online and the pace with which their volumes' back catalogs were made accessible on the internet,[46] so that, in general, online accessibility will differ between articles with different relevant past literature even if the articles were written in the same year.

2.3.1 Online accessibility treatment variable

In order to measure to which degree an article's relevant literature has been accessible online, we calculate the share of online accessible volumes[47] of all cited journals at the time the paper has (presumably) been drafted. More formally, we denote the set of all journals by $\boldsymbol{\mathcal{J}}$. Suppose an article i has been published in year t and cites the subset of journals $\boldsymbol{J}_i \subset \boldsymbol{\mathcal{J}}$. Indexing the journals in $\boldsymbol{J}_i$ by j, our online treatment is given by

$$T_i(t) = \frac{\sum_{j \in \boldsymbol{J}_i} a_j(t-1)}{\sum_{j \in \boldsymbol{J}_i} h_j(t-1)}, \tag{2.1}$$

where $a_j(t-1)$ is the number of volumes of journal j that have been accessible online in the year $t-1$ on at least one platform and $h_j(t-1)$ is the number of existing historical volumes at date $t-1$ published in journal j. This measure of online availability of relevant literature is article-specific, since it depends on the set of cited journals and their online accessibility.[48] In the empirical section the treatment defined in (2.1) is called *percent online*.

The default behavioral model behind our treatment variable is that an author facing zero online accessibility searches the relevant literature in print, for instance by browsing his library's collection of volumes aided by keyword or abstract indexation systems; in contrast, another author whose relevant literature is partially accessible online will browse this electronic literature by using internet tools, while still using the same methods as the previous author for the literature available only in print. In such a case, the use of internet literature browsing and searching tools coincides exactly with our online treatment variable. In practice, some deviations from such behavior are likely. For instance, very low levels of *percent online* might not induce researchers to search online; and, conversely, researchers whose literature is almost entirely online might neglect the few remaining print-only volumes. However, studies on researchers' literature searching behavior suggest that the joint use of print and electronic resources (with declining use of print) was typical for researchers during the transition to full electronic access (Tenopir, Hitchcock and Pillow, 2003; Boyce et al., 2004), so that *percent online* should be a reasonable approximation to researchers behavior.[49]

A qualification needs to be made at this point. While we can compute an article's share of relevant literature which was *accessible* online, we do not observe whether the article's authors effectively *used* the internet to search for related literature. Thus, online accessibility effects should be understood as intention-to-treat effects of online access. In section 2.5 we use information about aggregate time trends of subscriptions to online contents from other studies to explore the relationship between accessibility and access.

A second remark relates to the question whether *percent online* could be endogenously linked to diversity. Many plausible stories of endogeneity which rely on differences in journal specific time trends are excluded by our journal-year fixed effects specification. Another concern is that the top five journals were available online from very early on. Thus, articles citing predominantly top five journals might have a high measure of *percent online*. This would bias the estimated effect if such articles were inherently more/less diverse.

Similarly, researchers who tend to cite older literature (which has lower online accessibility, on average) might be inherently more/less diverse, too. Since any measure of online accessibility is bound to have such problems, we will address these concerns in the empirical section through appropriate robustness checks. For instance, we will test whether there is an online accessibility effect when comparing articles with the same share of top five journals cited or with the same age distribution of references.

A more challenging concern is that –beyond the differences between journals and over time– there might be some additional heterogeneity on the author level. Specifically, consider the hypothetical case where online accessibility has no effect on diversity, yet authors differ in the extent to which they make use of diverse sources. Then, if the propensity to adopt the internet is correlated with the diversity of an author's research agenda,[50] we would find spurious effects of the online treatment on diversity. We address this concern by including author fixed effects (in addition to the journal-year fixed effects). However, one should bear in mind that estimations with author fixed effects might be too conservative since they exclude channels through which the effect of online accessibility works. For instance, online access could change the composition of "diverse" and "non-diverse" authors. Therefore we see the regressions with author fixed effects as an important robustness check but not our main specification.

2.3.2 Online accessibility and diversity over time

Our estimation sample includes the 45,553 articles or paper-and-proceeding articles published between 1991-2009 in the 50 considered core journals. Figure 2.2 plots, for each year, the average share of existing volumes that were accessible online on at least one platform. Online accessibility of economics journals started in 1997 on the JSTOR platform. Since then, the back volumes of the different journals were gradually scanned and uploaded, and in 2009 almost all publications were available online. Hence, the considered period covers some years of the pre-internet era as well as the entire transition to full coverage. Until the turn of the millennium online access was clearly dominated by JSTOR. Later, other platforms caught up. A large amount of back volumes of Elsevier journals (which are not included in JSTOR) were made accessible in 2005.

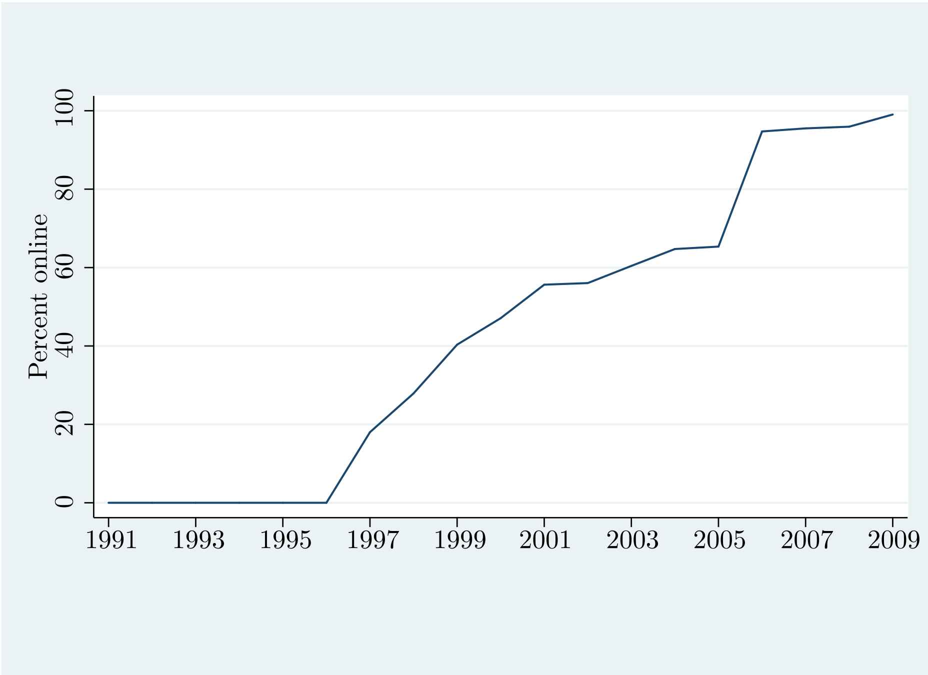


Figure 2.2: The share of economics journal volumes accessible online

Notes: The Figure plots for each year 1991-2009 the average share of existing volumes published in the 50 selected economics journals which was accessible online on at least one platform.

Figure 2.3 plots the average number of one-, two- and three-digit JEL codes assigned to an article. For the first years in our sample, the number of assigned codes is constant. Then, it rises for all JEL code digits from 1995 onwards. Figure 2.2 and 2.3 reveal a positive time correlation of online accessibility and the number of assigned JEL codes. But since the number of assigned JEL codes might not be comparable between different years we do not want to overstate this correlation. For instance, some additional JEL codes were added after 1991. Moreover, the assigning process might have changed over time.[51]

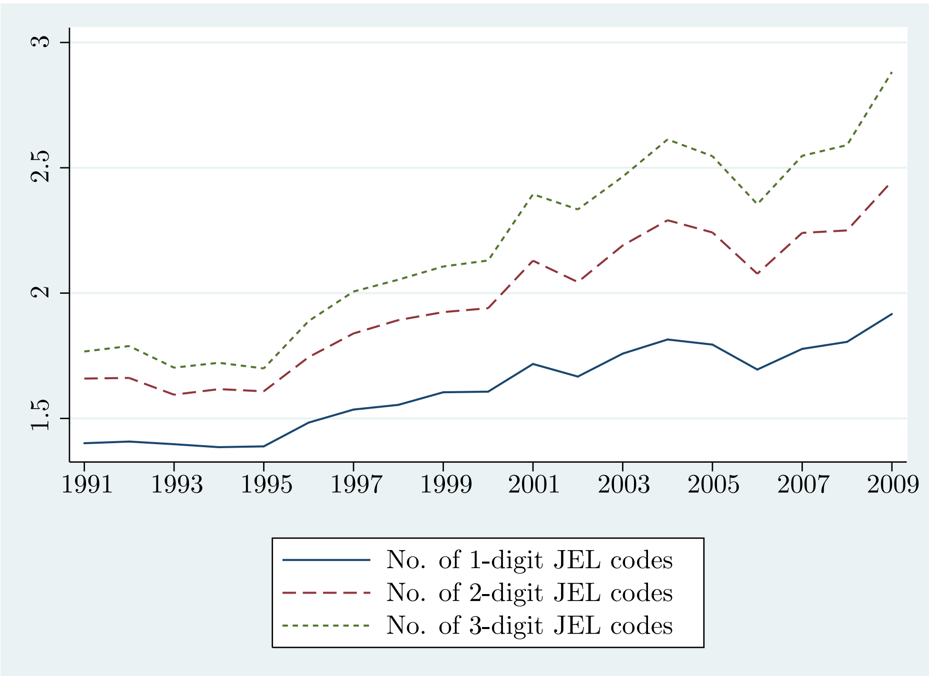


Figure 2.3: The average number of JEL codes over time

Notes: The figure plots the average number of one, two and three digit JEL codes. The sample includes the 45,553 articles published between 1991-2009 in the considered journals.

With 0.25, 0.27, 0.20 and 0.28, the four different categories of geodesic distances have about the same relative prevalence (the summary statistics of all variables can be found in Table B.8 in Appendix B.1). But these averages mask huge cross-sectional variations. In the case of the geodesic distances, the time trend is even harder to interpret than it is in the case of the number of assigned JEL codes. The way geodesic distances are constructed generates an inherent time trend, since the citation network is more comprehensive for later years where our dataset covers more back volumes. This makes the share of geodesic distances higher than three falling by construction.[52] It is the aim of our empirical strategy to exploit this cross-sectional variation while controlling for any secular time trends in order to estimate the effect of online access on the measures of diversity.

2.3.3 Identification strategy

As explained above, in the distribution of geodesic distances, time trends emerge by construction. Such inherent time trends are present in the assignment of JEL codes, too. For instance, some three- and two-digit codes, and even a one-digit code, have been introduced after 1991. Moreover, until the mid 1990's the production process set an upper bound of five for the number of codes assigned to an article. For all these reasons it is indispensable to control for year fixed effects to disentangle the causal effect of online accessibility from other ongoing trends.[53]

There are substantial differences in diversity measures between journals. While it is not obvious that this journal-specific diversity is related to online accessibility, such a correlation could arise if the relevant literature predominant in some journals was accessible online later or earlier than in others. To allow for changing time- and journal-specific heterogeneity in the most flexible way, we account for journal-year fixed effects. Thus, the variation which we use to estimate the effect stems from cross-article differences in the share of the relevant literature that is online within a given year and a given journal. Since the structure of the data consists of articles in journal-years forming an unbalanced pseudo-panel, we use the (linear) panel specification

$$Y_{iv} = \alpha T_{iv} + \boldsymbol{X}'_{iv}\boldsymbol{\beta} + \mu_v + u_{iv}, \tag{2.2}$$

where i indexes articles and $v = \tilde{v}(j,t)$ journal-years. Thus, Y_{iv} represents the diversity measure of article i which was published in journal j and year t. With slight abuse of notation, T_{iv} stands for the online treatment defined in (2.1). $\boldsymbol{X}_{iv}$ is a vector of possible article-specific control variables to be discussed below, and $\boldsymbol{\beta}$ is a conformable parameter vector. μ_v denote fixed effects specific to journals and years. Finally, u_{iv} is an idiosyncratic shock. Under mean independence of u_{iv} from T_{iv}, $\boldsymbol{X}_{iv}$, and μ_v, the coefficient α measures the causal effect of a marginal increase in the share of literature online on the diversity measure Y_{iv}. Equation (2.2) can be estimated conveniently using the OLS within-estimator.

2.4 The effect of online accessibility on diversity

2.4.1 Baseline results

Estimation results for the baseline model (2.2) are collected in Table 2.1. Panel I depicts the coefficient of the treatment variable *percent online* for regressions on the number of one-, two- and three-digit JEL codes (first three columns) and on the fraction of geodesic distances equal to one, two and three (last three columns). In all regressions, the panel dimension of the OLS within-estimator is journal-years, of which there are 859 unique groups. The standard errors shown are robust to heteroskedasticity and clustering at the journal-year level.

Since the average of *percent online* varies from zero in 1991 to one in 2009, the coefficients of the first three columns can be read as the total change in the average number of JEL codes comparing a world without any online access to

Table 2.1: Fixed effects regressions of percent online on diversity variables, N=45,553

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr(g=1)	Pr(g=2)	Pr(g=3)
I. *Regressions on percent online*						
Perc. online	0.2208***	0.3426***	0.3565***	-0.0670***	-0.0027	0.0589***
	(0.0274)	(0.0360)	(0.0408)	(0.0120)	(0.0104)	(0.0079)
R^2	0.0482	0.0679	0.0921	0.0044	0.0055	0.0219
II. *Regressions on percent online and further control variables*						
Perc. online	0.1792***	0.2823***	0.2955***	-0.0404***	-0.0344***	0.0356***
	(0.0265)	(0.0350)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
R^2	0.0727	0.0928	0.1135	0.1460	0.1018	0.0854

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables in Panel II: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

one which provides full access to all 50 journals. The effect is substantial: the coefficients —about 0.2 for one-digit JEL codes and about 0.35 for two- and three-digit JEL codes— correspond to 39, 45 and 32 percent of the increase of one-, two- and three-digit JEL codes in the data in the observed period.

Online accessibility has a diversity-enhancing effect on the distribution of geodesic distances, too: The fractions of low geodesic distances ($g = 1, 2$) are reduced and higher geodesic distances ($g = 3$, $g > 3$) are increased as a consequence of online accessibility. The coefficients -0.067, -0.0027, 0.0589 and 0.0108, respectively can be read as percentage-point changes in the fractions of geodesic distances.[54] Thus, the fraction of shortest geodesic distances (whose average over the entire period is 25 percent) is estimated to have shrunken by about 6.7 percentage points due to online accessibility of the literature.

While the regressions in Panel I control for any confounding journal-year-specific characteristics, there might be some further heterogeneity within journal-years correlated to the treatment *percent online* which could drive the effect. This is a stronger concern for the regressions on geodesic distances. For instance, higher shares of long geodesic distances can result from citing two types of relatively unconnected work. The first possibility is that the references in question, while well-connected to other literatures, are relatively unconnected between them. The second possibility is that the references in question are relatively unconnected at all. While we would readily interpret the first case as a sign of diversity, some might want to exclude the second case from counting as diversity. To address this issue, Panel II adds to the specification the average number of citations received by a reference and the average number of references contained in a reference. In this way, the online accessibility effect is computed for similarly well-connected reference networks. The list of other control variables in Panel II includes a papers-and-proceedings dummy, the number of authors, number of pages, number of references made, number of distinct journals referenced, percent references found in the data, and percent of self-references.

The effects in Panel II remain large and statistically significant. The coefficients for the JEL code regressions are somewhat smaller than before. This is mainly the result of controlling for number of pages and number of different journals referenced, two variables mediating the effect of online accessibility on diversity. Whether these variables are part of the causal effect and should not be controlled for is to a large extent a matter of taste and interpretation. The effect on the distribution of geodesic distances is slightly larger overall, with about 7.5 percentage points being shifted from the lower part of the distribution ($g = 1, 2$) to the right tail ($g = 3$, $g > 3$). While without controls the shift was mainly from $g = 1$ to $g = 3$, now the effect is more evenly distributed among the four categories of g.[55]

2.4.2 Robustness checks

An important first robustness check for our results relates to the appropriate lag of the treatment. The time when an article's references were collected is unknown and has to be inferred from the date of publication; there is also bound to be differences in length of the publication process across articles. In (2.1) we made the informed guess that the best approximation is the online accessibility faced by authors one year prior to publication.[56] Table B.2 in Appendix B.1 explores alternative lags of zero, two and three years. Given the heterogeneity in publication process length, it would be worrisome to find that the results in Table 2.1 hold only under the one year lag. Comfortingly, the results remain qualitatively the same for all lags explored, although the effects are strongest for the one- and two-year lag, which is in line with our expectations that a majority of the articles' time from draft to publication lies in the one-to-two-year range.

Next, we set out to assess the robustness of our treatment by exploring other ways of capturing online accessibility. Implicitly, the treatment *percent online* gives more weight to long-standing journals (with many volumes) because the percentage is calculated over the sum of all cited journals' volumes. An alternative which weights journals equally is to construct the treatment as the percent online in the average journal cited.[57] Similarly, treatment can be defined as the percent of an article's references that were online one year prior to publication. This weights the journals by their share in the reference list. Finally, instead of focusing on percentages, treatments can also be constructed based on the *absolute number* of volumes online (an approach related to Evans, 2008). Table B.3 in Appendix B.1 documents that the baseline results from Table 2.1 remain valid for any of these alternative treatments.

Another robustness check is with respect to the data sources of online access of the different platforms. Note that our measure of online accessibility combines information obtained directly from JSTOR with the one collected by FSO. Detailedness and quality of these two sources varies. Whereas FSO collects its data twice a year, JSTOR's database is very precise.[58] To make sure that such differences between data sources are not influencing our results, we constructed two treatments: one taking into account access provided by JSTOR only, the second taking into account access on the remaining online platforms contained in the FSO data. The results (displayed in Table B.6 in Appendix B.1) show that disaggregating the treatment by data source delivers estimates that are very similar to the aggregate treatment in the baseline specification.

The OLS estimator used in Table 2.1 gives the best linear fit for our model of diversity and online accessibility without relying on strong distributional assumptions. It also has the attractive property that the effects for the geodesic distances add up to zero. In Table B.4 in Appendix B.1 we explore an alternative, constant-elasticity specification which we estimate both by OLS (using the logarithm of the dependent variables) and by Poisson Pseudo-Maximum Likelihood. Again, the effects (which are now to be interpreted as approximate percental changes) are similar.

The last issue we explore is a refinement of the fixed effects. While defining fixed effects at a more detailed level can purge the online accessibility effect from more confounding through unobserved heterogeneity, there is a trade-off to be considered since such an approach entails a loss of precision because of the higher number of fixed effects. In a first step, we treated papers-and-proceedings issues of a journal as a separate journal. Since most journals publish such issues, the number of panel units for these regressions increased to $1,456$. As a logical consequence of such an approach we can go even one step further and define a separate fixed effect for every single issue published in every journal in the period. This gives over 4,800 fixed effects. With both specifications, as the estimates in Table B.5 in Appendix B.1 show, the results are only marginally affected.

2.4.3 Author fixed effects

A more fundamental refinement of the fixed effects changes the panel dimension to a much less aggregated unit: the authors. While in our baseline regressions we exploit the variation in online accessibility between articles of a particular journal in a given year, a different source of variation comes from repeated publications of the same authors.

Exploiting only the variation for a given author (group) changes the interpretation of the coefficients, as the online accessibility effect being estimated excludes some channels which are part of the effect using the within journal-year variation. For instance, the availability of online literature may have an impact on the composition of the pool of authors, increasing the share of authors which are efficient users of online tools. In the estimation with journal-years fixed effects this margin is part of the causal effect as the pool of authors is not kept constant and changes with the spread of online accessibility. While ultimately we favor this approach, the specification with author fixed effects provides an important alternative view which shows the effect of online accessibility for authors publishing repeatedly during this period.

We approach this issue from two perspectives. In the first take, we extract from the EconLit database all author names which appear at least in two articles, leaving us with $12,165$ distinct authors.[59] We cloned articles with multiple authors to create one record for every author, and obtained a total of $67,903$ observations. Observations corresponding to the same article are clearly not independent, and the reported standard errors account for this correlation. Indeed, we used two-way clustered standard errors (Cameron, Gelbach and Miller, 2011) which are robust to heteroskedasticity and clustering at the article level, as well as at the author level. Moreover, in addition to the control variables, we included a full set of journal-year indicator variables to account for these fixed effects, too. The results are printed in Panel I of Table 2.2. Panel II contains results where the panel units are co-author-groups (including groups of size one, i.e. single authors). There are $7,307$ such unique co-author-groups which have published more than twice in our

Table 2.2: Author and co-author-groups fixed effects

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
I. *Regressions with author fixed effects,* $N = 67,903$						
Perc. online	0.0606**	0.1409***	0.1473***	-0.0213*	-0.0277**	0.0329***
	(0.0300)	(0.0385)	(0.0420)	(0.0123)	(0.0108)	(0.0092)
R^2	0.1617	0.1893	0.2283	0.1887	0.1676	0.1118
II. *Regressions with author-group fixed effects,* $N = 21,767$						
Perc. online	0.0512	0.1068*	0.1612**	-0.0009	-0.0165	0.0242
	(0.0418)	(0.0571)	(0.0640)	(0.0188)	(0.0157)	(0.0148)
R^2	0.1044	0.1404	0.1789	0.1475	0.1031	0.0685

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator. Regressions in Panel I account for author fixed effects (12,165 groups). Panel I standard errors (in parentheses) robust to heteroskedasticity and clustering at author (12,165 groups) and article level (41,441 groups). Regressions in Panel II account for co-author-group fixed effects (7,307 groups). Panel II standard errors (in parentheses) robust to heteroskedasticity and clustering at co-author-group level. Further control variables in both Panels: complete set of journal-year indicators, paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

data. The total number of articles they have written is $21,767$. As before, we additionally include over 800 journal-year fixed effects and our list of control variables. Standard errors are robust to clustering at the co-author-group level.

The results in Table 2.2 are substantially less precise. Given the substantial loss of degrees of freedom, this does not come as a surprise. It is the more remarkable, therefore, that the results in this table reveal the same patterns than those from the baseline regressions. To be sure, the coefficients are attenuated compared to the baseline; still, we find that online accessibility significantly increased the number of JEL codes, and that it transferred probability mass from the lower end of geodesic distances' distribution ($g = 1, 2$) to its right tail.

2.5 Heterogeneity of the effects and further results

Having established the robustness of the effect of online accessibility on the diversity of academic articles in economics, this section explores the heterogeneity of the effect and possible channels through which it affects the diversity measures.

2.5.1 The effect over time

A potential source of heterogeneity in the effect is time. While there are many potential factors with a time trend, one of them has been highlighted in the literature as particularly relevant for online accessibility: institutional subscription to platforms providing online contents of economics journals (i.e. effective online access). For instance, Depken and Ward (2009) and McCabe and Snyder (2011) document that the number of institutions subscribing to JSTOR (and to Elsevier's online contents) increased in the considered period almost linearly (cf. Depken and Ward, 2009, Fig. 1, McCabe and Snyder, 2011, Fig. 7). Table 2.3 shows estimation results for a specification which adds an interaction of the treatment with a linear time trend, which is bound to capture this effect of increasing online access. The time trend was normalized to zero in 1997, so that the coefficient on *percent online* gives the effect in that year. The effect on the JEL codes is indeed moderate in the beginning of the period and shows an increasing time trend, suggesting that as more researchers gained access to online contents the effect of online access on the number of JEL codes became more prominent. However, results are less clear-cut for the geodesic distance regressions, where the absence of a time trend cannot be rejected.

Table 2.3: Treatment interacted with time trend, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr(g=1)	Pr(g=2)	Pr(g=3)
Percent online	0.0366	0.1101*	0.0941	-0.0479**	-0.0265	0.0195
	(0.0484)	(0.0635)	(0.0696)	(0.0213)	(0.0162)	(0.0141)
Perc. online × year	0.0319***	0.0385***	0.0450***	0.0017	-0.0018	0.0036
	(0.0088)	(0.0122)	(0.0142)	(0.0044)	(0.0034)	(0.0030)
F-statistic	32.05	36.85	29.98	6.72	5.78	11.11
p-value	0.0000	0.0000	0.0000	0.0013	0.0032	0.0000
R^2	0.0705	0.0948	0.1247	0.1483	0.0966	0.0827

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. The variable "Perc. online × year" is normalized to zero in 1997. F-statistics and p-values are for joint significance tests on coefficients of "Percent online" and "Perc. online × year". R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

2.5.2 The effect across journals

The effect of online accessibility found in our baseline regressions could vary greatly for different journals. While for the average journal the effect on diversity is positive, it could be that this aggregation masks negative effects for certain classes of journals. To explore this issue we estimated a specification with interactions for three classes of journals: the "top five" journals, general interest journals, and field journals (Table 2.4).[60] We find the same kinds of effects as in the baseline regressions for every journal category. The effects are most pronounced for second tier general interest journals, but the effects are large for all three categories.

2.5.3 The composition of referenced journals and publication years

One way in which online accessibility may have influenced diversity is by reducing the bibliographic importance of the journal an article appeared in. The correlation between reading a particular journal and contributing to it might have been weakened by the internet, leading to a more diverse pool of influences. A second way in which online accessibility may have influenced diversity is by increasing the importance of the "top five" journals, which are journals publishing diversely to begin with. Two characteristics of these journals are that they have a long publication history and that they were among the first to be put online. Thus, researchers relying on online sources were likely to rely on these journals. Table 2.5 addresses this issue. The specifications for which results are shown add two regressors to the model: the percent of an article's references that were published in the journal where the article appeared, and the percent of an article's references that were published in one of the "top five" journals. The results indicate that these two channels explain some of the online accessibility effect. The variables themselves are highly significant, the percent references to "top five" journals increasing diversity, the percent references to the own journal decreasing it.

Table 2.4: Treatment interacted with journal type, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr(g=1)	Pr(g=2)	Pr(g=3)
Perc. online × top 5	0.1519***	0.2344***	0.2780***	-0.0239	-0.0498**	0.0170
	(0.0458)	(0.0577)	(0.0687)	(0.0214)	(0.0227)	(0.0133)
Perc. online × gen. interest	0.2613***	0.4154***	0.3856***	-0.0608**	-0.0311*	0.0391***
	(0.0673)	(0.1015)	(0.1102)	(0.0246)	(0.0184)	(0.0129)
Perc. online × field	0.1727***	0.2739***	0.2813***	-0.0443***	-0.0266**	0.0452***
	(0.0375)	(0.0481)	(0.0542)	(0.0149)	(0.0122)	(0.0107)
R^2	0.0699	0.0907	0.1130	0.1458	0.1021	0.0829

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

The online accessibility effect on JEL codes is reduced by about 30 to 50 percent. It remains significant, however, suggesting that there are further channels at work as well. The pattern is similar but less accentuated for geodesic distances, where the joint reduction in the fraction of distances one and two is about 20 percent.

Finally, we set out to quantify the importance of the age distribution of an article's reference for the effect on diversity. The results from regressions including average citation lag (i.e. the difference in years between the article's publication year and that of its average reference) are shown in Table 2.6. Average citation lag has been used as the primary dependent variable in previous work analyzing the impact of online accessibility on academic research (Evans, 2008 and Depken and Ward, 2009).[61] The fact that our coefficients of interest remain virtually unaffected in size and statistical significance when including citation lag shows that our direct measures of diversity go substantially beyond the heterogeneity of references' age distribution.

2.6 Concluding remarks

This paper documents how online accessibility of articles lead to an increase in the diversity of upcoming economic research. We do so by considering local measures of diversity, i.e. the diversity of ideas *a single article* touches on or is based on. This is a sharp contrast to the aggregate measures of diversity considered in Evans (2008) or McCabe and Snyder (2011). It can well be, that the local diversity increases at the same time as the number of overall cited articles decreases and the concentration of cited articles increases (as suggested by Evans, 2008). For instance, different fields of economics may get tighter connected, whereas in each field some "superstars" emerge. However, the results of McCabe and Snyder (2011) suggest that in the case of economics, online access did not skew the distribution of citations and did lead to a decline in the number of uncited papers. Whether local or aggregate measures of diversity are of interest depends on the context. And

Table 2.5: Citing top 5 journals and own journal, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr(g=1)	Pr(g=2)	Pr(g=3)
Percent online	0.0928***	0.1742***	0.2183***	-0.0461***	-0.0096	0.0310***
	(0.0293)	(0.0380)	(0.0428)	(0.0115)	(0.0106)	(0.0079)
Percent refs. to top 5	0.2478***	0.3589***	0.2448***	0.0431***	-0.0982***	0.0137
	(0.0420)	(0.0529)	(0.0569)	(0.0119)	(0.0101)	(0.0085)
Perc. refs. to own journal	-0.2895***	-0.1569**	-0.1597**	0.0929***	-0.0313**	-0.0136
	(0.0465)	(0.0609)	(0.0663)	(0.0158)	(0.0128)	(0.0107)
R^2	0.0728	0.0929	0.1090	0.1432	0.1218	0.0859

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 2.6: Average citation lag, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	$\Pr(g=1)$	$\Pr(g=2)$	$\Pr(g=3)$
Percent online	0.1820***	0.2878***	0.3052***	-0.0396***	-0.0302***	0.0380***
	(0.0266)	(0.0351)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
Average citation lag	-0.0019***	-0.0037***	-0.0067***	-0.0005**	-0.0029***	-0.0016***
	(0.0007)	(0.0009)	(0.0011)	(0.0002)	(0.0002)	(0.0002)
R^2	0.0732	0.0937	0.1154	0.1453	0.1075	0.0917

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

it is unclear whether more diversity is a priori desirable. In the introduction, we provided one example of a setting where such local diversity matters and is desirable – a model of scientific research based on recombinant growth and limited attention. In any case, our results suggest that online access did not narrow but broaden the mind of economic researchers.

Notes

30. This dependence on the stock of preexisting knowledge is often called the "standing on the shoulders of giants" effect (which goes back to a quote by Isaac Newton and now serves as an advertising slogan of Google Scholar).

31. For a theoretical contribution how diversity can be measured and ranked see Weitzman (1992) and (1998b). Stirling (2007) emphasizes "variety", "balance" and "disparity" as three district properties of diversity. Consisting of these three components, Van den Bergh (2008) analyzes optimal diversity in a model of recombinant innovation.

32. A much larger literature exists on the impact of access to internet contents in more traditional market settings. In a recent review of this literature, Brynjolfsson, Hu and Smith (2010) stress that a key channel through which information technology improvements changed how consumers learn about goods and services, and how producers develop, distribute and deliver them, is through a transformation in search and recommendation tools.

33. It is important to emphasize that these JEL codes are not the ones declared by the authors of a paper. The codes we use are assigned by a team of economists at EconLit.

34. American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics and Review of Economic Studies.

35. The set of journals includes all journals considered in the standard Tilburg ranking, as well as the list considered in Palacious-Huerta and Volij (2004). Furthermore, it includes all core journals in Conroy et al. (1995), all journals used in Kalaitzidakis, Mamuneas and Stengos (2003), as well as all top 20 journals in Combes and Linnemer (2011). The list is comparable to Depken and Ward's (2009) and McCabe and Snyder's (2011) who include 79 and 63 economics journals, respectively. Using eigenfactor.org's list of over 200 economics journals, we found that our list has an eigenfactor score of around 0.75 in 1995. I.e., randomly traversing the citation network spanned by all economics journals, the list's 50 journals are selected 75 percent of the time.

36. Bavelas (1948) and Freeman (1979) provide early foundations. See Newman (2003) for an overview.

37. A related strand of the literature on scholarly communication studies social networks defined by co-autorship relationships (Newman 2001a,b,c), as opposed to information networks defined by citations. This literature, too, is foremost descriptive. For an exception, see Kretschmer (2004) who links co-author networks features to author productivity.

38. The analysis of patents is also conceptually related to scholarly communication networks. For instance, social networks akin to co-authorships are defined by inventor collaborations, and information networks arise in the context of patent citations. Recent examples of papers studying geodesic distances are Balconi, Breschi and Lissoni (2004), in the former context; and Lee, Su and Wu (2010), in the latter.

39. "The Fed and Interest Rates: A High-Frequency Identification," *American Economic Review, Papers and Proceedings*, May 2002, Volume 92, Issue 2, pp. 90-95.

40. In the following we abbreviate all the sources by authors and publication date in italic without specifying the entire reference. For the exact reference of the citations we refer the reader to the paper by John Cochrane and Monica Piazzesi.

41. See Pencavel (1991), the editor's note with which the Journal of Economic Literature introduced the new system. The JEL codes replaced an earlier, more narrow classification system.

42. JEL codes have been the subject of some descriptive work which used them to characterize the evolution of economic fields or subfields over time (Kim, Morse, Zingales, 2006; Kelly and Bruestle, 2011). Previous literature using JEL codes in regression analysis has included them as control variables for the specific fields (e.g. Formby, Gunther and Sakano, 1993, Axarloglou and Theoharakis, 2003, Boschini and Sjögren, 2007).

43. We assume that no volumes were accessible before 1998 on platforms covered by the FSO data. This is reasonable, since only about 2 percent of volumes were online on platforms other than JSTOR in 1998, and these accessible volumes were mainly the contemporaneous ones. For the historically most important platform of online access, JSTOR, we do have the data about the accessibility of journals even prior to 1998.

44. JSTOR is not included in the FSO database before 2009.

45. Evans (2008), whose units of observation are journal-years, considers the journal where an article is published as the only relevant past literature. This approach seems unsuitable for economics. In our data, on average only 7 percent of citations refer to the same journal where the article is published. Even for the journal where this ratio is highest over the whole period – the Journal of Finance – it is only 20 percent.

46. Cf. Evans (2008), Depken and Ward (2009), McCabe and Snyder (2011).

47. We use the term "volume" to denote all issues of a journal published in the same calendar year.

48. The online treatment defined in (2.1) is based on two specific assumptions regarding *(i)* the set of relevant literature and *(ii)* the time lag between first draft and publication of a paper. In section 2.4 we show that our results are robust to other reasonable specifications.

49. An alternative interpretation of the online treatment variable can be given assuming a different, more stylized behavioral model where there exist only two types of researchers: one group using print literature only, the other group relying exclusively on online literature. Then, *percent online* can be interpreted as the probability that the article's author is an online researcher.

50. For instance, younger researcher might adopt the new technology faster and might differ from their older colleagues in terms of the diversity of their research interests.

51. E.g. the dint in the number of codes in 2006 might be explained by a change of EconLit's managing director. The assigning process is, however, consistent within a given year.

52. It is possible to try and capture this by partialling out some time trend. However, this requires assumptions for the trend's functional form. In Figure B.1 in Appendix B.1 a possible correction has been applied. In that graph the average fractions of short geodesic distances (order one and two) fall over time, while those of order higher than three increase. This would be in line with an increase in diversity over the period. In contrast to Figure B.1, our regression framework presented in the next section, which does include time fixed effects, does not require arbitrary assumptions about the time trend.

53. See also McCabe and Snyder (2011) who illustrate the empirical importance of flexible controls of time trends.

54. The coefficient of *percent online* in a regression on $\Pr(g > 3)$ is not shown in the table, but it can be easily obtained from the three numbers depicted in the last three rows of the table, since it is equal to the negative of the sum of the coefficients in the regressions for $g = 1, 2, 3$ (because shifts in the probability mass of the distribution add up to zero).

55. The full set of estimates is shown in Table B.1 in Appendix B.1; summary statistics for dependent variables and all regressors are in Table B.8 in Appendix B.1.

56. Evans (2008) uses the same lag specification, while Depken and Ward (2009) and McCabe and Snyder (2011) use a lag of zero.

57. Formally, the treatment is then calculated as $\frac{1}{|\boldsymbol{J}_i|}\sum_{j\in\boldsymbol{J}_i}\frac{a_j(t-1)}{h_j(t-1)}$ instead of (2.1).

58. In fact, we know from JSTOR for each journal issue the exact date of first user access.

59. We used data from EconLit as we found it significantly more reliable than Thomson Reuters', which contained numerous inconsistencies in the coding of author names.

60. Note that uninteracted level effects are subsumed in the journal-year fixed effects.

61. Table B.7 contains further regressions including the median and standard deviation of references' publication year. These results lead to the same conclusion that our measures of diversity capture a different dimension of heterogeneity than the distribution of citation lags.

Chapter 3

How does mechanization and structural change sum up to a constant aggregate labor's income share?

Chapter Summary

For decades, theories of economic growth aimed to replicate a constant aggregate labor's income share – for empirical as well as technical reasons. However, some sectors feature a systematically declining labor's share whereas it is constant in others. Consequently, the observed constancy of the aggregate factor shares is only due to a counterbalancing effect of structural change (i.e. labor intensive sectors account for an increasing fraction of the aggregate economy). Standard growth models do not account for this complexity. This paper provides a stylized growth theory in which both dynamics – structural change and mechanization on a sectoral level – are present and endogenous technical change leads to a constant aggregate labor's income share.

3.1 Introduction

The labor's income share is a widely debated variable with a long tradition in the economic literature. Often it served as a proxy for the income distribution and involved political brisance. Nowadays it has become a key (calibration) parameter in macroeconomic models in general. The dynamic of the labor's income share is importantly linked to technical change, growth and development.

In the first half of the 20th century there emerged a consensus in the economic literature that the labor's income share is on the aggregate level constant over time. Kaldor (1961) postulated the constancy of labor's share to be one of the "stylized facts" of growth. And since then, Kaldor's stylized characterization remains a good approximation of the data. Figure 3.1 plots the U.S. labor's income share for the period 1960-2007.[62] Apart from smaller fluctuations over the business cycle the variable is indeed very stable around a value of 0.623.

However, it is not only due to this empirical fact, that virtually any theoretical growth model aims to replicate constant factor shares. The steady state concept of growth is indispensably connected to a constant labor's share.[63] Therefore, the existence of a balanced growth path (and the solvability of dynamic models more generally) depends on a, at least asymptotically, constancy of factor shares.

There are two well-known and frequently used specifications which allow to replicate a constant labor's income share in an one-sector theory of economic growth: A Cobb-Douglas technology (see Cobb and Douglas, 1928) or exclusively labor augmenting technical progress which occurs at the same rate as the accumulation of capital. The vast majority of endogenous growth models chooses the first solution with an Cobb-Douglas technology at the micro-level.[64,65] However, empirical evidence – as well as common-sense – speaks for an elasticity of substitution between labor and capital which is strictly smaller than unity.[66]

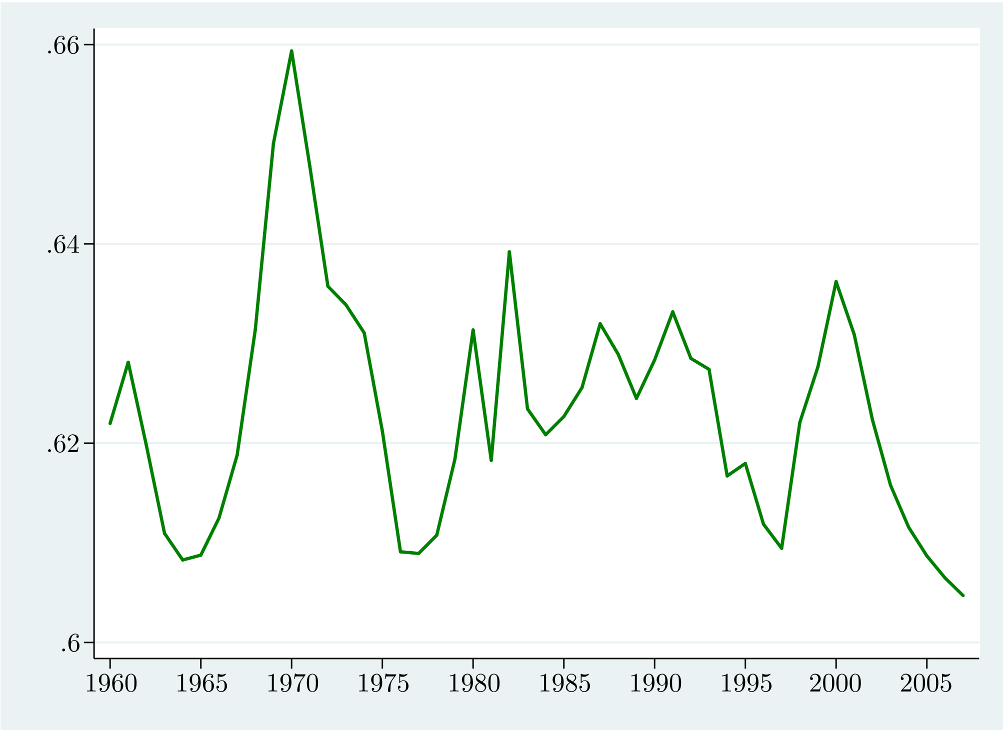


Figure 3.1: Aggregate labor's income share in the U.S.

Notes: The figure plots the labor's shares of income in in the U.S. for the years 1960-2007. The labor's share is calculated as total labor compensation relative to total value-added. Source: World KLEMS (November 2011 release).

Jones (2005) is a rare example of a one-sector endogenous growth theory which can replicate a constant labor's share although working with a production function where capital and labor are complements. The model turns out to be consistent with a constant aggregate labor's share since from the perspective of a used micro-level production function, the technical progress is purely labor augmenting.[67] Another related contribution is Acemoglu (2003) which shows that a model of directed technical change is consistent with purely labor augmenting technical change if capital and labor are gross complements and there is "extreme state dependence", which means that there are no knowledge spillovers between the capital and labor augmenting knowledge stock.

This paper illustrates that the constancy of the labor's income share is due to the joint evolution of the labor's share at sectoral level and structural change. Consequently, in order to understand the mechanism behind the constancy of factor shares we need a multi-sector model with sectoral technologies which allow for changing labor's income shares (i.e. which are not Cobb-Douglas). Figure 3.2 depicts the evolution of the labor's income share in three broadly defined sectors: a "progressive sector", which encloses sectors as agriculture, mining, manufacturing, transportation and communication; a "non-progressive sector" which mainly summarizes service categories and a "neutral" sector which consists of real estate.[68] We see that the labor's share of the non-progressive and neutral sectors are constant over time whereas the one of the progressive sector is declining. In this paper, the decline in the progressive sector's labor share is called "mechanization".[69]

How can Figure 3.2 be consistent with a constant aggregate labor's share? Note that by definition, the aggregate labor's income share at date t, $\alpha(t)$, is a weighted sum of the sectoral labor's shares $\alpha_j(t)$, i.e.

$$\alpha(t) = \sum_{j=B,N,U} \alpha_j(t) S_j(t), \tag{3.1}$$

where the weights, $S_j(t)$, $j = B, N, U$ are the value-added shares of the pro-

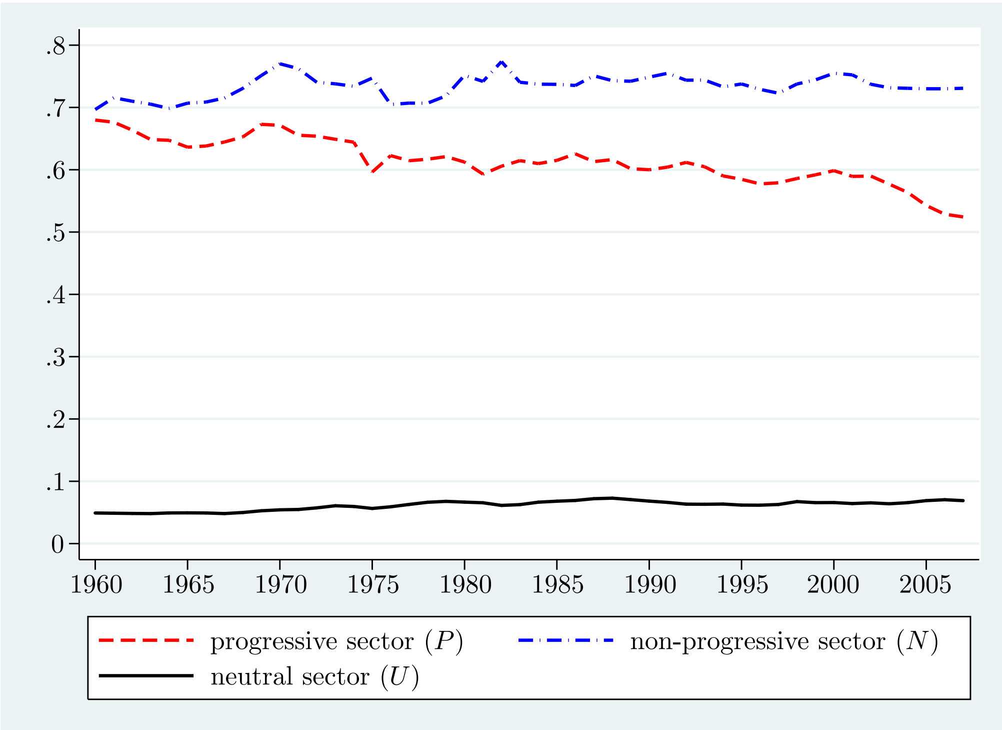


Figure 3.2: Labor's income shares on the sectoral level

Notes: The figure plots the labor's income shares of three broadly defined sectors in the U.S. for the years 1960-2007. The labor's shares are calculated as total labor compensation relative to total value-added. The "progressive sector" comprises the ISIC sections A to E and I (Agriculture, Hunting, Forestry and Fishing; Mining and Quarrying; Total Manufacturing; Electricity, Gas and Water Supply; Transport and Storage and Communication), whereas the "non-progressive sector" contains the sections F to H and J to Q except K70 (Construction; Wholesale and Retail Trade; Hotels and Restaurants; Finance, Insurance and and Business Services; Community Social and Personal Services). The "neutral sector" consists of K70 (i.e. "Real estate activities"). Source: World KLEMS (November 2011 release).

gressive, non-progressive and neutral sector, respectively.[70] For the aggregate factor shares to be constant the change in sectoral value-added shares (i.e. the structural change) must exactly compensate the decline in the progressive sector's labor income share. This is possible if sectors with a relatively high labor's share constitute a continuously increasing fraction of the GDP.

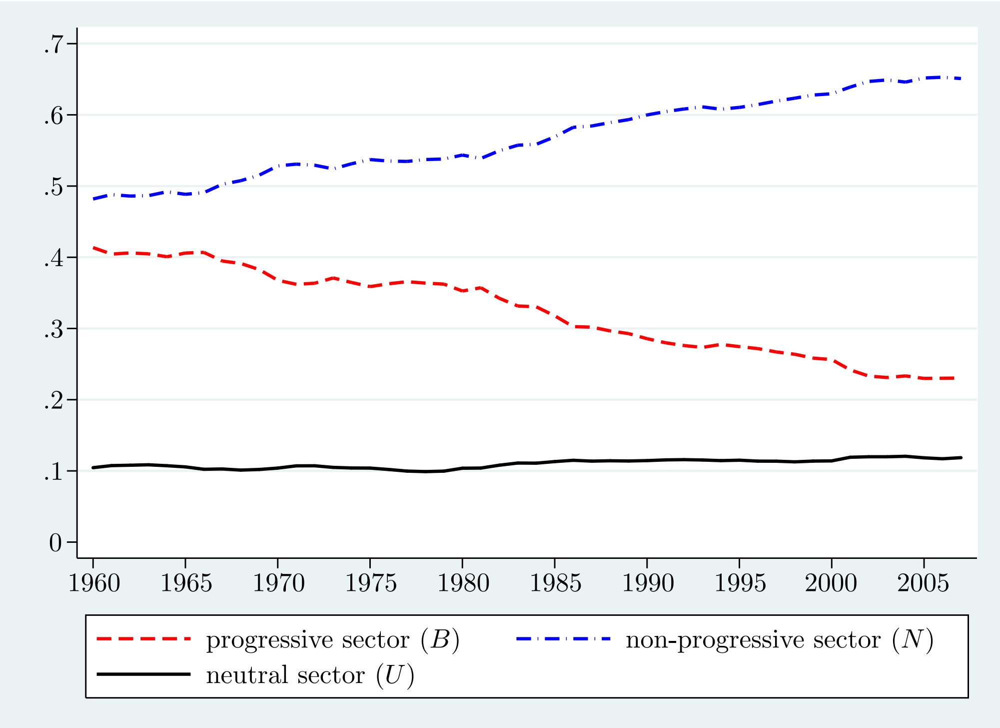


Figure 3.3: Value-added shares on the sectoral level

Notes: The figure plots the value-added shares of income in three broadly defined sectors in the U.S. for the years 1960-2007. The "progressive sector" comprises the ISIC sections A to E and I (Agriculture, Hunting, Forestry and Fishing; Mining and Quarrying; Total Manufacturing; Electricity, Gas and Water Supply; Transport and Storage and Communication), whereas the "non-progressive sector" contains the sections F to H and J to Q except K70 (Construction; Wholesale and Retail Trade; Hotels and Restaurants; Finance, Insurance and and Business Services; Community Social and Personal Services). The "neutral sector" consists of K70 (i.e. "Real estate activities"). Source: World KLEMS (November 2011 release).

Figure 3.3 shows the evolution of the value-added shares of the three sectors. Indeed, the relatively labor intensive non-progressive sector is clearly expanding at a faster rate than the aggregate economy, whereas the progressive sector's value-added share is declining. The neutral sector is characterized by a constant value-added share. So by taking the first derivative of (3.1) with respect to time we see that the observed constancy of the aggregate labor's share (i.e. $\alpha(t) = \alpha$) is the product of the counterbalancing forces:

$$\dot{\alpha}_B(t) = \frac{\dot{S}_B(t)}{S_B(t)} \left[\alpha_N - \alpha_B(t)\right], \tag{3.2}$$

where we already made use of the (empirical) observation that the α_N, α_U and S_U are approximately constant over time. This equation (3.2), which ties together the structural change between the progressive and non-progressive sector and the pace of mechanization in the progressive sector, highlights the underlying mechanism behind constancy of the aggregate factor share.[71] Note that since both the labor's share and the value-added share of the neutral sector are constant this third sector is important in order to get the aggregate *level* of the labor's income share right, but is dispensable for studying its *dynamics*.

It is the aim of this paper to provide a theory in which equation (3.2) is endogenously fulfilled. But why should the mechanization of the progressive sector and the structural change comove in such a particular way? In the provided theory, directed technical change constitutes a link between the two variables. On the one hand, directed technical change affects the evolution of relative prices between sectors and therefore the structural change (as long as the elasticity of substitution is not equal to unity). And on the other hand, the ratio between the interest and wage rate determines the incentive to engage in productivity enhancing activity in the progressive sector. This investment incentive ultimately leads to a constant aggregate labor's share.[72]

The remainder of the paper is organized as follows: Section 3.2 provides the theoretical model and Section 3.3 discusses the results and concludes.

3.2 Theoretical model

3.2.1 Representative household

There is a representative infinitely living agent with the following indirect preferences

$$\mathcal{U}(0) = \int_0^\infty \exp\left[-\rho t\right] \log\left[U\left(E(t), P_B(t), P_N(t)\right)\right] dt, \tag{3.3}$$

where $U\left(E(t), P_B(t), P_N(t)\right)$ is linearly homogeneous in the nominal expenditures $E(t)$ and declining in the prices of the progressive and non-progressive sectors $P_B(t)$ and $P_N(t)$. We have $\rho > 0$. The intratemporal budget constraint is given by

$$X_B(t)P_B(t) + X_N(t)P_N(t) = E(t), \tag{3.4}$$

where $X_j(t)$, $j = N, B$ represents consumed output of the progressive and non-progressive sectors.

Assuming logarithmic intertemporal preferences is crucial in order to reconcile balanced growth with structural change caused by relative sectoral price dynamics (see Ngai and Pissarides, 2007). In line with the structural change literature we assume that the utility function implies an elasticity of substitution between the two sectors which is strictly smaller than unity. The representative household owns $V(t)$ units of wealth in a given instant of time and supplies inelastically L units of labor. $V(0)$ is exogenously given. The households maximize utility subject to a standard flow budget constraint

$$\dot{V}(t) = V(t)r(t) + Lw(t) - E(t),\ \forall t, \tag{3.5}$$

and a transversality condition

$$\lim_{t\to\infty}\left[V(t)\exp\left[-\int_0^t r(s)ds\right]\right] \geq 0. \tag{3.6}$$

Intertemporal optimization then leads to the following Euler equation

$$g_E(t) = r(t) - \rho, \tag{3.7}$$

where $g_E(t)$ is the growth rate of nominal expenditures. Intratemporal optimization allows us to write the optimal expenditure share devoted to the progressive sector, $S_B(t)$, as a decreasing function of the relative price $\frac{P_B(t)}{P_N(t)} \equiv p_B(t)$, i.e.

$$S_B(t) \equiv \frac{P_B(t) X_B(t)}{E(t)} = \tilde{S}_B\left(p_B(t)\right), \tag{3.8}$$

with $\frac{\partial \tilde{S}_B(p_B(t))}{\partial p_B(t)} < 0$. The fact that $S_B(t)$ is decreasing in $p_B(t)$ is due to the assumption that the elasticity of substitution is strictly smaller than unity.

3.2.2 Production and capital accumulation

Suppose the output of the progressive sector is produced competitively with the following technology (see Aghion and Howitt, 2005)

$$X_B(t) = \epsilon^{-\epsilon}(1-\epsilon)^{-(1-\epsilon)} \int_0^{M(t)} A_i(t)^{1-\epsilon} x_i(t)^{\epsilon} \left[\frac{L_B(t)}{M(t)}\right]^{1-\epsilon} di. \tag{3.9}$$

This is a Cobb-Douglas function defined over a CES aggregate of intermediate inputs $x_i(t)$, $i \in [0, M(t)]$ and labor $L_B(t)$. We have $\epsilon \in (0,1)$. $A_i(t)$ is the "quality" parameter of intermediate input variety i. Each variety i is produced by a monopolist with constant marginal cost

$$c_i(t) = c\left(\frac{w(t)}{D(t)^{\Delta}}, r(t)\right), \ \forall i, \tag{3.10}$$

where $D(t)^{\Delta}$ is a labor-augmenting spillover, with $\Delta > 1$. Let us assume that the technology (3.10) features an elasticity of substitution which is less than unity. Then, the labor's income share in the production of variety i is increasing if and only if $\frac{w(t)}{C(t)^{\Delta}}$ increases relative to $r(t)$. Note that in neoclassical growth theory (as well as in the data) $w(t)$ increases relative to $r(t)$. Hence, in order to reconcile a declining labor's share with a micro production function with gross complements we need technical change, i.e. an increasing $D(t)$.

The production function of the non-progressive sector is linear in labor

$$X_N(t) = L_N(t). \tag{3.11}$$

Besides the two consumption goods, physical capital can be accumulated according to the following law of motion

$$\dot{K}(t) = I(t), \tag{3.12}$$

where $I(t)$ are investments (in terms of the investment good). The investment good is produced according to a linear production function in capital

$$I(t) = AK_I(t), \tag{3.13}$$

where $K_I(t)$ is capital allocated to the investment good sector. We have $A > \rho$.

Except of the variety markets $i \in [0, M(t)]$ all markets are perfectly competitive. Labor and capital is perfectly mobile across sectors and its factor reward is denoted by $w(t)$ and $r(t)$, respectively. In the following we choose the investment good as a numéraire. We then obtain the following lemma.

LEMMA 3.1. *Optimal price setting and cost minimization implies*

$$p_B(t) = \left[\frac{c\left(\frac{w(t)}{D(t)^{\Delta}}, r(t)\right)}{(1-\epsilon)w(t)}\right]^{\epsilon} \bar{A}^{-(1-\epsilon)}, \tag{3.14}$$

where $\bar{A} \equiv \frac{1}{M}\int_0^M A_i di$. *The choice of numéraire implies*

$$A = r(t), \forall t. \tag{3.15}$$

Proof. Progressive output producers solve the following optimization problem (suppressing time indices):

$$\min_{\{x_i\}_{i=0}^{M}, L_B} wL_B + \int_0^M p_i x_i di$$

subject to

$$X_B = \epsilon^{-\epsilon}(1-\epsilon)^{-(1-\epsilon)} \int_0^M A_i^{1-\epsilon} x_i^{\epsilon} \left[\frac{L_B}{M}\right]^{1-\epsilon} di,$$

where X_B is an exogenous output level. The multiplier attached to the constraint can be interpreted as the (competitive) price of the output of the progressive sector P_B. Solving the optimization problem we get

$$P_B = \left[\frac{Mw}{\int_0^M A_i p_i^{-\frac{\epsilon}{1-\epsilon}} di}\right]^{1-\epsilon}, \tag{3.16}$$

and for the demand for x_i

$$x_i = \frac{\epsilon}{1-\epsilon} A_i p_i^{-\frac{1}{1-\epsilon}} \frac{wL_B}{\int_0^M A_i p_i^{-\frac{\epsilon}{1-\epsilon}} di}. \tag{3.17}$$

Given this iso-elastic demand, it is optimal for each monopolist i to charge a constant mark-up, i.e. $p_i = \frac{c(\cdot)}{\epsilon}$. Substituting this into (3.16) we get

$$P_B = \left[\frac{c(\cdot)}{\epsilon}\right]^{\epsilon} w^{1-\epsilon} \bar{A}^{-(1-\epsilon)}, \tag{3.18}$$

where $\bar{A} \equiv \frac{1}{M}\int_0^M A_i di$. Moreover, since we have $P_N = w$ the relative price can be written as (3.14). Finally, (3.15) follows immediately from the choice of numéraire. Q.E.D.

Because of the iso-elastic demand the monopolists set a constant mark-up over marginal costs and consequently all prices are proportional to marginal costs. Then, the relative price p_B is determined by two components: First, by the ratio of the marginal costs of an intermediate input producer i and of a final non-progressive output producer. And second by the average quality of intermediate inputs. Note that the relative marginal costs depend on the relative factor intensities and the relative factor price $\frac{w(t)}{r(t)}$. Since the production function of the non-progressive sector is linear in labor we have two additional facts: (i) The labor's share of the non-progressive sector is constant over time and (ii) the non-progressive sector is *always* more labor intensive than the progressive sector.[73] Both of these implications are consistent with the empirical patterns observed in Figure 3.2. Fact (ii) implies that p_B is increasing in $r(t)$ and decreasing in $w(t)$.

The fact that the investment good denominated interest rate is constant over time is simply due to the AK technology in the investment good sector.

I make two additional assumptions: First, the spillover term is proportional to the capital stock, i.e.

$$D(t) = D_0 K(t), \tag{3.19}$$

where D_0 is a constant. Second, as in Aghion and Howitt (2005) section 5, in order to destroy a scale effect we assume that $M(t)$ expands proportionally to $L_B(t)$, or formally[74]

$$M(t) = L_B(t). \tag{3.20}$$

3.2.3 Quality choice

In every point in time t, each of the monopolists has to choose its quality level $A_i(t)$. In order to produce quality $A_i(t)$ at date t the monopolist has to incur fixed cost of

$$\left[\frac{A_i(t)}{\bar{A}(t)}\right]^{1+\phi} D(t)r(t). \tag{3.21}$$

The term in squared brackets states that it is more expensive to have a quality which exceeds the average quality level. This term ensures that all monopolists choose the same quality level at a given point in time. $D(t)$ is a (negative) spillover term. As the economy gets more "complex" (i.e. as $D(t)$

increases) more capital is needed to run a production process of the same quality.

Monopolists are exposed to a continuous threat to lose their monopoly position. In any instant of time an incumbent is overthrown if a potential entrant is willing to produce at a higher quality. This implies that monopolists make zero profits in equilibrium, or $A_i(t)$ must be such that

$$x_i(t)\left[p_i - c(\cdot)\right] = \left[\frac{A_i(t)}{\bar{A}(t)}\right]^{1+\phi} D(t)r(t). \tag{3.22}$$

The left-hand side of this equation are profits before fixed costs, whereas the right-hand side are fixed costs. By substituting in the equilibrium demands and prices we obtain the following lemma.

LEMMA 3.2. *The optimal quality choice of monopolist i is characterized by*

$$\left[\frac{\epsilon}{D_0}\right] w(t) = \left[\frac{A_i(t)}{\bar{A}(t)}\right]^{\phi} K(t)r(t). \tag{3.23}$$

Proof. (3.23) is obtained by substituting (3.17), $p_i = \frac{c(\cdot)}{\epsilon}$ and (3.19) into (3.22). Q.E.D.

A direct implication of (3.23) is that all intermediate input producer i choose the same quality level, $A_i(t) = \bar{A}(t)$, $\forall t$. Then, (3.23) reduces to

$$\Xi w(t) = K(t)r(t), \tag{3.24}$$

with $\Xi = \frac{\epsilon}{D_0}$. Now since L is constant over time and there are no other production factors than capital and labor,[75] (3.24) directly implies that the *aggregate* labor's income share is constant. Any other situation is inconsistent with the optimal quality choice. Why is this the case? First, the exogenous change in the number of available varieties, (3.20), implies that profits before fixed costs are proportional to the wage rate. Second, the assumption that quality investments are in terms of capital makes the cost of quality improvement linear in the interest rate. Consequently, the optimal quality choice is a trade-off which involves the relative factor price $\frac{w(t)}{r(t)}$. Finally, the spillover which is proportional to the size of the economy (see (3.19)) makes the factor shares constant. However, another characteristic of the cost of quality improvement specification (3.21) is also important: The chosen quality level enters the equation only relative to the average. Hence we have positive spillovers and from a social point of view a higher $\bar{A}$ does not require more physical investments. This allows $\bar{A}$ to adjust flexibly until condition (3.24) is fulfilled.

Knowing that the $\bar{A}(t)$ adjusts such that the aggregate labor's share is constant allows us to characterize the aggregate equilibrium dynamics. But before first we have to summarize the market clearing conditions.

3.2.4 Market clearing conditions

Since physical capital is the only production factor which can be accumulated asset market clearing implies

$$V(t) = K(t), \ \forall t. \tag{3.25}$$

Labor and capital market clearing is given by

$$K(t) = K_I(t) + \int_0^{M(t)} k_i di + \int_0^{M(t)} k_i^A di, \tag{3.26}$$

and

$$L = L_B(t) + L_N(t) \int_0^{M(t)} l_i(t) di, \tag{3.27}$$

where $k_i(t)$ and $l_i(t)$ denote capital and labor employed by monopolist i and k_i^A is capital used for quality investments by monopolist i. Finally, goods market clearing can be expressed as

$$S_B(t)E(t) = P_B(t)X_B(t) \tag{3.28}$$

and

$$[1 - S_B(t)]\, E(t) = P_N(t) X_N(t). \tag{3.29}$$

3.2.5 Dynamic equilibrium

We define a dynamic equilibrium as a path of prices and quantities which fulfills household and firm optimization. We obtain the following proposition:

PROPOSITION 3.1. *There is an unique dynamic equilibrium path in which household expenditures, the stock of physical capital and the wage rate grow at identical, constant rate*

$$g_E(t) = g_K(t) = g_w(t) = A - \rho > 0, \ \forall t. \tag{3.30}$$

The labor's income share and the saving rate are constant and equal to

$$\alpha(t) = \frac{L}{L + \Xi} \equiv \alpha, \forall t. \tag{3.31}$$

and

$$s(t) = \frac{A - \rho}{A}(1 - \alpha), \ \forall t. \tag{3.32}$$

The investment good denominated interest rate is constant and equal to A.

Proof. The Euler equation, the asset market clearing condition, (3.24) and the flow budget constraint imply

$$\dot{K}(t) = A \left[1 + \frac{L}{\Xi}\right] K(t) - E(0) \exp\left[(A - \rho)t\right]. \tag{3.33}$$

This differential equation has the following solution

$$K(t) = Z \exp\left[\left[1 + \frac{L}{\Xi}\right] At\right] + \frac{E(0)}{\frac{AL}{\Xi} + \rho} \exp\left[(A - \rho)t\right], \tag{3.34}$$

where Z is a constant to be determined. Substituting this into the transversality condition gives

$$\lim_{t\to\infty}\left[Z\exp\left[\frac{AL}{\Xi}t\right]+\exp\left[-\rho t\right]\frac{E(0)}{\frac{AL}{\Xi}+\rho}\right]=0. \tag{3.35}$$

since $\frac{AL}{\Xi}>0$ we must have $Z=0$. This implies that $\frac{\dot{K}(t)}{K(t)}=\frac{\dot{w}(t)}{w(t)}=\frac{\dot{E}(t)}{E(t)}=A-\rho$, $\forall t$, where the first equality comes from (3.24). The labor's income share is constant and given by (3.31) due to (3.24), too. The saving rate can be rewritten as

$$s(t)=\frac{AK_I(t)}{w(t)L+AK(t)}=\frac{K_I(t)}{K(t)}(1-\alpha). \tag{3.36}$$

Note that $\frac{K_I(t)}{K(t)}=\frac{A-\rho}{A}$ in order that the capital stock grows at the balanced growth rate. Q.E.D.

It is remarkable that we can explicitly solve for the aggregate variables along the dynamic equilibrium path. There is no transitional dynamic and the economy's aggregate behavior is fully disconnected from the evolution of the sectoral variables. This is first due to the logarithmic intertemporal preferences which imply that the saving rate is constant although the relative price p_B changes. Second it is due to (3.24) which guarantees that the factor shares are constant no matter how things change at the sectoral level.

The evolution of the sectoral variables depends on the functional form of $\tilde{S}_B(p_B)$ and $c(\cdot)$. Because production of the non-progressive sector is linear in labor the labor's income share is unity in this sector, i.e. $\alpha_N=1$. If we spend one unit of money for output of the progressive sector a fraction of

$$\alpha_B(t)=1-\epsilon+\epsilon^2\alpha_{vc}(t), \tag{3.37}$$

will be used for labor compensation, where $\alpha_{vc}(t)$ is the labor's share implied by the variable cost (3.10).[76]

Note that both $w(t)$ and $D(t)$ grow at constant rate $A - \rho$ whereas $r(t)$ is constant over time. As a result, since $\Delta > 1$, $\frac{w(t)}{D(t)^{\Delta}}$ declines relative to $r(t)$. Hence if (3.10) implies that labor and capital are gross complements, $\alpha_{vc}(t)$ is declining over time. This generates a picture which is consistent with the empirical motivation (i.e. Figure 3.2 and 3.3) in the introduction.

3.3 Critical and concluding remarks

Section 3.2 provided a theoretical framework with a constant aggregate labor's income share generated by offsetting trends in value-added and factor shares at a sectoral level. However, I have to acknowledge that this result depends on several specific assumptions. In the following the key assumptions are critically analyzed. The discussion is also aimed to summarize potential challenges for future research.

The core piece of the theory is equation (3.23) which states that firms have an incentive to change their level of sector specific quality investments whenever the aggregate labor's income share deviates from a constant α. In order to get this result we had to overcome several issues. One challenge is that in an environment of structural change a sector specific innovation incentive changes due to changing market size. In this paper I solve this issue by the specification of Aghion and Howitt (2005), section 5 (see (3.20)). This implies that the innovation incentive is proportional to the wage rate although the sectoral market shares change over time. Although Aghion and Howitt's (2005) specification is somewhat ad hoc it should be possible to endogenize it by a so called "second generation growth model" where we have both expanding variety as well as quality improvement.

Another critical assumption is the spillover term $D(t)$ in (3.21) which has to be proportional to $K(t)$. However, with an endogenously expanding set of product varieties – as suggested above – there might be a way to rationalize this spillover.

A clearly very unsatisfactory element is the assumption of "perfect" spillover in equation (3.21) such that only the relative quality level $\frac{A_i(t)}{A(t)}$ affects the costs.[77] Relaxing this assumption should definitely be the aim of future research.

Other criticism might be that although the mechanization of the progressive sector is emphasized in the introduction, the mechanization in the model occurs only due to a spillover and is not driven by a specific incentive.

This leads us to the conclusion that the stylized model presented in Section 3.2 is not the complete answer to everything. There are many directions in which the theory can and should be improved. But hopefully, if this is done, such a theory allows us to learn more about the essence of the aggregate labor's share than a theory which just assumes an aggregate Cobb-Douglas production function with $\alpha = 0.623$.

Notes

62. Note that the labor compensation calculations of the KLEMS takes the labor income of proprietors into account – in contrast to the BEA data (see Krueger, 1999). This is done by assuming that, in a given industry, the proprietor's hourly compensation is the same as the average worker's compensation. See also Gollin (2002) for the importance of this adjustment.

63. With any constant return to scale production function and competitive factor rewards the labor's share can be written as $1 - \frac{f'(k)k}{f(k)}$ where k is the capital stock per efficiency units of labor. Hence, in order to have a stationary point in k, the labor's share must be constant too (and vice versa).

64. Many even assume that the micro-level technology is linear in labor, while generating a constant fraction of capital income via a constant mark-up.

65. Clearly, if all production factors can be accumulated and grow at identical rate, balanced growth and constancy of factor shares can be reconciled. But such a theory would imply constant relative factor prices between human and physical capital and no physical capital deepening, which is clearly at odds with the data.

66. See footnote 3 of Acemoglu (2003) and the papers cited therein.

67. In Jones' (2005) model, at each point in time there exists a number of independent draws of a capital and labor requirement. The production possibilities in a point in time are then the hull of all drawn combinations of input requirements. It is shown that if the input coefficients are drawn from two independent Pareto distributions the hull converges to a Cobb-Douglas technology (as the number of draws goes to infinity).

68. Other papers may refer to the first two sectors as the "good" and the "service" sector. But since the first sector is associated with a relatively high total factor productivity (TFP) growth and consequently

declining prices relative to the aggregate price index I prefer the labels "progressive"/"non-progressive" (which go back to Baumol, 1967).

69. Many papers document a declining labor's income share in the manufacturing sector (see e.g. Böckerman and Maliranta, 2012). Some papers relate this decline to foreign trade competition. In this paper however, we will focus on a closed economy and explain the dynamics by technical change. In fact, the decline in Figure 3.2 is very monotonic over the last 50 years and does not seem to accelerate with globalization. And Figure C.1 in the Appendix illustrates that a declining labor's share in the progressive sector is a pattern we can observe in many countries.

70. By definition we have $\sum_{j=B,N,U} S_j(t) = 1$.

71. The two counteracting forces are quantitatively important. The mean value-added share of the progressive sector is 0.32 and the progressive sector's labor share declines in the 47 years by about 16 percentage points. Hence abstracting from the counteracting structural change the aggregate labor's share would decline by over 5 percentage points in the observed period.

72. This is a big contrast to the existing literature about structural change. All papers of structural change with a constant aggregate labor's share I am aware of generate the constancy by identical and constant labor's shares across sectors. For structural change in terms of value-added, Figure 3.2 clearly rejects this assumption. Moreover, note that in order to get a constant aggregate labor's share with structural change we need mechanization. With structural change and (constant) capital intensity differences across sectors as in Acemoglu and Guerrieri (2008) aggregate factor share constancy is not possible.

73. In general factor intensity reversal can occur if we have two sectors with different elasticities of substitution.

74. Note that in contrast to Aghion and Howitt (2005), where the scale effect is generated by population growth changing market size arises in this model due to structural change (i.e. changes in $S_B(t)$).

75. Note that firms make in equilibrium zero profits, such that there is no value of firm ownership.

76. Note that (3.9) is a Cobb-Douglas defined over a CES aggregate and labor, where $1-\epsilon$ is the weight on labor. Hence a fraction $1-\epsilon$ is directly used for labor compensation. The other fraction ϵ is used to cover the variable and fixed costs. Fixed cost are entirely used for capital compensation, whereas a fraction $\alpha_{vc}(t)$ of the variable costs goes to the workers. Because of the constant mark-up the weights of variable and fixed cost are given by ϵ and $1-\epsilon$, respectively.

77. This assumption somewhat resembles the one of "perfect state dependence" of the knowledge accumulation in Acemoglu (2003).